Analysis and Design of Line of Sight MIMO transmission systems

POLITECNICO DI TORINO

SCUOLA DI DOTTORATO

PhD Course in
Ingegneria Elettronica e delle Comunicazioni – XXVI cycle

PhD Dissertation

Analysis and Design of Line of Sight MIMO transmission systems

Christian Camarda

Supervisor
Prof. Guido Montorsi

PhD Director
Prof. Ivo Montrosset

March 2014

Contents

List of Figures

Chapter 1

Introduction

1.1 Background and objective

A cost-effective solution to the problem of guaranteeing backhaul connectivity in mobile cellular networks is the use of point-to-point microwave links in the Q-Band and E-Band [3, 4, 5]. The always increasing rate in mobile data traffic makes these microwave radio links a potential bottleneck in the deployment of high-throughput cellular networks. This consideration has stimulated a large body of research aimed at the design of high-capacity backhaul links [6, 7, 8, 3]. One design challenge is that the use of high-order constellations to increase throughput (512 QAM has been recently demonstrated in commercial products) makes the overall system extremely sensitive to phase noise. Another example are communication systems employing low-cost low-quality RF oscillators, such as in DVB-S2 transceivers (see [9] and references therein) and in the large-MIMO transceivers currently under theoretical investigation [10].

A fundamental way to characterize the impact of phase noise on the throughput of these systems is to study their Shannon capacity. Unfortunately, the capacity of the phase-noise channel is not known in closed-form, even for simple channel models.

The effect of phase noise in telecommunication systems is more evident in presence of multiple antennas at transmitter and receiver because of the overlapping of phase noise contribution in receivers. For this reason the second fundamental challenge nowadays is well-recovering phase oscillations in order to achieve good performance.

The thesis proposes in Chapter 2 a simulated-based tool to compute a lower bound to channel capacity for SISO and MIMO systems in presence of phase noise with one oscillator shared among the antennas per side. In Chapter 3 we give a non-asymptotic expression of an upper bound to capacity always for SISO and MIMO channels and finally in Chapter 4 we show a low complex phase detector based on a

combination of Phase Locked Loop (PLL) exploiting the decisions made by a turbo decoder.

The aim of this work is showing a way to bound the channel capacity for single-antenna and multiple-antennas channels impaired by phase noise generated by instabilities in oscillators driving all the transceivers, and compare the performance of the proposed phase detector to those theoretical limits.

A huge part of the work behind this thesis is related to a research project made to improve the performance of an actual backhaul link provided by the company Ericsson and to get a reference in terms of channel capacity.

1.2 Phase noise

Phase noise, caused by both phase and frequency instabilities in the radio-frequency (RF) oscillators used in wireless transceivers, is one of the major impairments in certain communication systems [11]. In this Section we present three different way to model the phase noise process: the well known Wiener model, the characterization of phase noise by mean of a mask realized by channel measurement campaign or simulation and, finally, phase noise can be described by the overlapping of first order Auto-Regressive (AR1) processes.

Having an accurate model describing the evolution in time and frequency of phase noise is essential to represents the channel by a thorough input-output relationship between the source and the receiver of a wireless backhaul link.

1.2.1 Wiener model

Phase noise can be modeled as a Wiener process [9, 12, 11], which is a random process defined as having zero mean normally distributed phase increments $[\theta(t_2) - \theta(t_1)]$ over any interval $[t_1, t_2]$ and independent phase increments over disjoint time intervals. We assume an incremental variance over a signaling interval equal to σ_Δ^2 and also assume that the channel phase $\theta(t)$ is slowly varying such that it can be considered constant over a symbol time T. In other words we assume that only samples of $\theta(t)$ at discrete-time nT are significant. These samples satisfy the discrete-time Wiener model, defined as[1]

$$\theta_k = \theta_{k-1} + \Delta_k, \quad k = 0, \ldots, n \tag{1.1}$$

where $\{\Delta_k\}$ s a process made by real, independent and identically distributed wrapped Gaussian random variables with zero mean and standard deviation σ_Δ.

[1] See [13] for a discussion on the limitations of this model.

The i.i.d. assumption on $\{\Delta_k\}$ implies that $\{\theta_k\}$ is a Markov process. Specifically,

$$f_{\theta_k|\theta_{k-1},\ldots,\theta_0} = f_{\theta_k|\theta_{k-1}} = f_\Delta \tag{1.2}$$

where

$$f_\Delta(\delta) \triangleq \sum_{l=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \exp\left\{-\frac{(\delta - 2\pi l)^2}{2\sigma_\Delta^2}\right\}, \quad \delta \in [0,2\pi)\,. \tag{1.3}$$

In other words, f_Δ is the probability density function of the innovation Δ_k modulo 2π.

Under the additonal assumption that θ_0 is assumed uniformly distributed in $[0,2\pi)$, we can state that the process $\{\theta_k\}$ is stationary.

The differential entropy of the Markov process in (1.2) is

$$h\left(\{\theta_k\}\right) = h(\Delta) \leq \frac{1}{2}\log(2\pi e\sigma_\Delta^2). \tag{1.4}$$

The upper bound, which holds because the Gaussian distribution maximizes the differential entropy under variance constraint [14, Theorem 8.65], turns out to be tight whenever $\sigma_\Delta \lesssim 50°$, as shown in [2], and, under this consraint, the expression of wrapped Gaussian pdf in (1.3) can be approximated to a standard Normal pdf.

1.2.2 Phase noise mask

To characterize an actual channel can be useful making measurement campaign to study the behavior of phase noise components at different frequencies. In Figure 1.1 and Figure 1.2 are depicted the power spectral density (PSD) of phase noise measured by Ericsson in an actual backhaul link respectively for 64-QAM and 256-QAM. In Chapter 4 we take those masks as phase noise reference to show the performance of a phase detector suitable for real scenarios.

To characterize the phase noise masks we indicate the reference phase noise power level at frequency $f = 10^5$ Hz. In Figure 1.1 the reference power level is -81 dB/Hz for 64-QAM, while for 256-QAM in Figure 1.2, due to the higher density of constallation, we experience a reference level of -87 dB/Hz.

1.2.3 Auto-regressive model

We now consider a phase noise model exploiting first order Auto-Regressive Gaussian processes (AR1) as presented in [12]. As we showed in Section 1.2.2 a phase noise mask, provided by actual channel measurements, is a realistic scenario for phase noise model. By analyzing, for example, the phase noise mask depicted in Figure 1.1, we note that it can be seen as the sum of two different slopes, corresponding to the behavior of phase noise at low frequency and at high frequency, and hence the

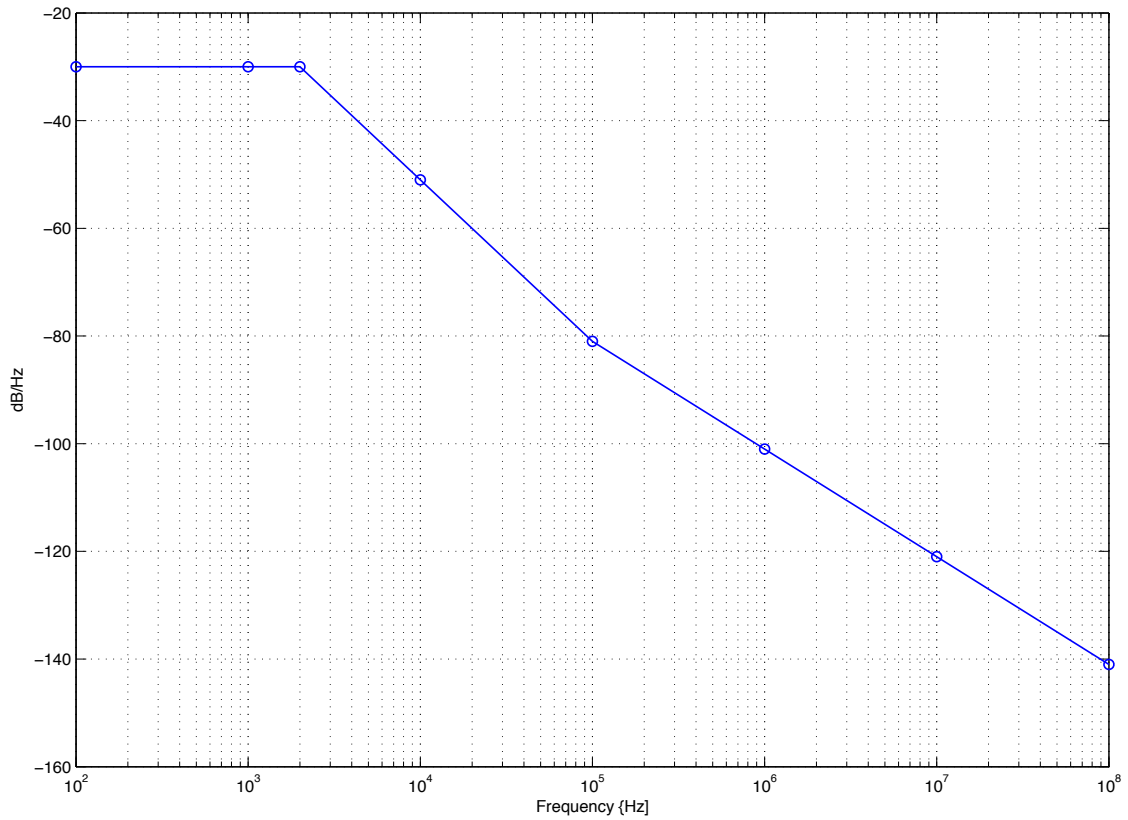

Figure 1.1. Power spectral density mask of measured phase noise provided by Ericsson for 64-QAM

phase noise process can be well represented by the sum of two AR1 processes. In detail, one AR1 describes the phase noise behavior at low frequency (slow process characterized by high power) while the second one represents the high frequency variations of phase noise (fast process characterized by lower power). In general a Gaussian AR1 process is defined as

$$u_k = a u_{k-1} + v_k \tag{1.5}$$

where v_k is a zero-mean Gaussian random variable, with variance σ^2 and where a is a real value such that $|a| < 1$ to ensure stability. By defining T_s the sampling period, the power spectral density (PSD) of u_k is

$$S_u(f) = \frac{\sigma^2}{1 + a^2 + 2a\cos(2\pi f T_s)}, \quad f \in \left[-\frac{1}{2T_s}, +\frac{1}{2T_s}\right]. \tag{1.6}$$

Thus, considering the sum of two AR1 process, we have

$$\theta_k = u_{a,k} + u_{b,k} \tag{1.7}$$

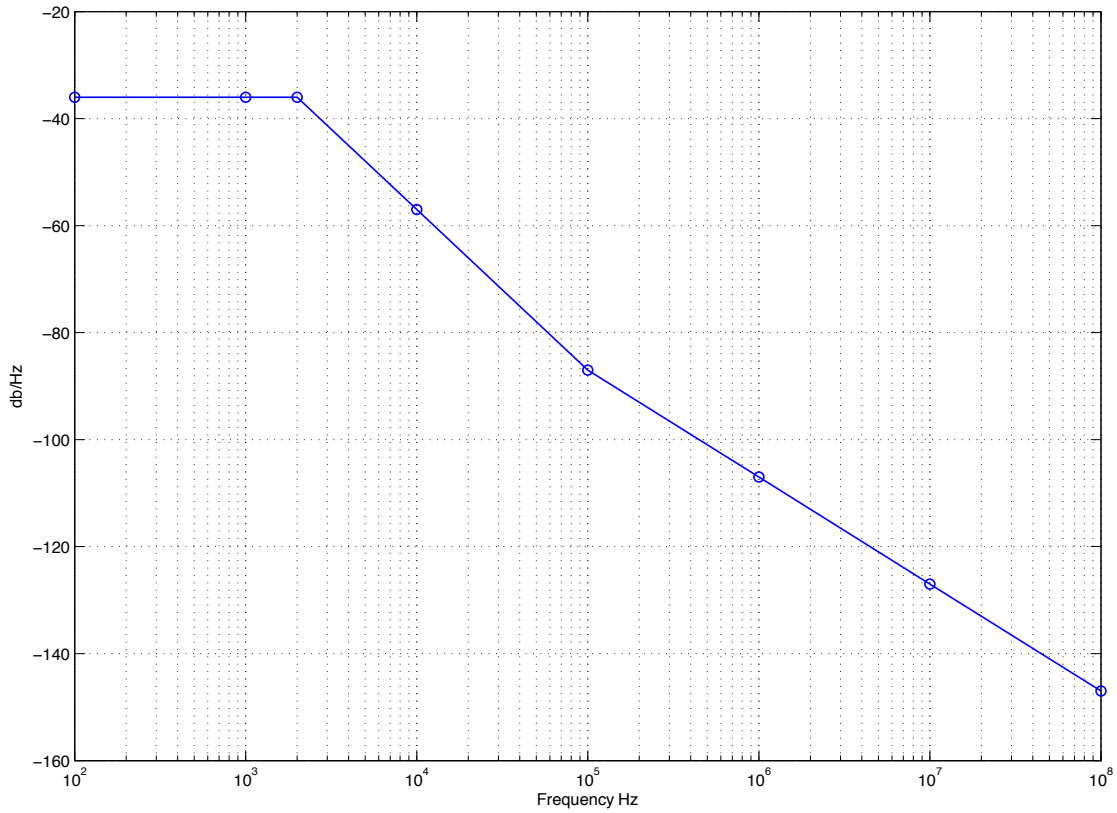

Figure 1.2. Power spectral density mask of measured phase noise provided by Ericsson for 256-QAM

where

$$u_{a,k} = au_{a,k-1} + v_{a,k} \tag{1.8}$$
$$u_{b,k} = bu_{b,k-1} + v_{b,k}. \tag{1.9}$$

Finally the resulting PSD of (1.7) is

$$S_\theta(f) = \frac{\sigma_a^2}{1 + a^2 + 2a\cos(2\pi f T_s)} + \frac{\sigma_b^2}{1 + b^2 + 2b\cos(2\pi f T_s)}, \tag{1.10}$$
$$f \in \left[-\frac{1}{2T_s}, +\frac{1}{2T_s}\right].$$

Hence we can approximate a realistic representation of phase noise, provided by a phase noise mask, with a double-AR1 process finding the four parameters $a, b, \sigma_a^2, \sigma_b^2$ such that the target phase noise mask is well fitted by (1.10).

1.3 MIMO Line of Sight

In the recent years, the ever increasing request for fast wireless communications has urged to design systems with high-throughput Line of Sight (LoS) backhaul links [14]. To obtain the desired efficiencies, two main engineering solutions have been resorted to: multiantenna links and high-efficiency modulations. In comparison with single-antenna links, MIMO LoS systems offer a relevant throughput increase but, at the same time, are more sensitive to phase noise, especially when design considerations impose to have different oscillators feeding each antenna at both sides.

The idea behind $M \times M$ MIMO LoS is to achieve a full-rank channel matrix $\mathbf{H}$ over a LoS link by a careful placement of the antennas at the transceivers [15, 16, 17]. Indeed, when the antenna spacing d at the transmitter and the receiver satisfies

$$d \approx \sqrt{\frac{\lambda R}{M}} \tag{1.11}$$

where λ is the wavelength and R denotes the distance between the transmitter and the receiver, the channel matrix $\mathbf{H}$ can be made not only full-rank, but also unitary [16, 17].

We next discuss some implications of (1.11) on the design of microwave backhaul links. Consider, for instance, a microwave backhaul link operating in the E-Band at 80 GHz. Assume that the transceivers are equipped with 2 antennas each and are 500 m apart. According to (1.11), the antenna spacing that results in a unitary channel matrix is about 97 cm, which is compatible with the assumption of using a single oscillator to drive the RF circuitries of both antennas.

In some cases, it may be convenient to locate the two antennas closer than what (1.11) prescribes. Then, $\mathbf{H}$ ceases to be unitary, although it can still be made full rank [17].

For a microwave backhaul link operating at 20 GHz over a 3 Km link, (1.11) results in an antenna spacing of about 3.8 m, which calls for a distributed oscillator solution.

In this thesis we will focus exclusively on the single-oscillator scenario and we will consider only the case of $\mathbf{H}$ unitary, with shared oscillators. The channel matrix $\mathbf{H}$ we use in next Chapters is

$$\mathbf{H} = \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}. \tag{1.12}$$

This model is accurate for MIMO systems where the distance between the antennas at the transceivers is sufficiently small for the RF circuitries at each antenna to be driven by the same oscillator [3].

Chapter 2

Lower bound to capacity

In this Chapter we describe a simulation-based technique for the computation of the mutual information (there is no maximization over all channel input distributions) by a recursive algorithm for SISO and MIMO system impaired by phase noise. Capacity lower bounds obtained by numerically computing the information rates achievable with various families of finite-cardinality independent and identically distributed (i.i.d.) input processes (e.g., QAM, PSK, and APSK constellations) have been reported in [9, 1, 18]. The numerical evaluation of these bounds is based on the algorithm for the computation of the information rates for finite-state channels proposed in [19].

Mutual information $I(\mathbf{X};\mathbf{Y})$ quantifies the amount of information that can be carried on a channel with input process $\mathbf{X}$ and output process $\mathbf{Y}$, expressed in bits per channel use. In the following work we will focus on the case where both $\mathbf{X}$ and $\mathbf{Y}$ are discrete-time stationary sequences denoted as $\mathbf{x}$ and $\mathbf{y}$ respectively. Mutual information between $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$I(\mathbf{x};\mathbf{y}) \triangleq \lim_{n\to\infty} \frac{1}{n} I(x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n). \tag{2.1}$$

From information theory [14] we know that, for each channel, $I(\mathbf{x};\mathbf{y})$ can be expressed as

$$I(\mathbf{x};\mathbf{y}) \triangleq H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}) - H(\mathbf{x},\mathbf{y}) \tag{2.2}$$

where $H(\mathbf{x})$ is the entropy rate of a source generating the random discrete-time process $\mathbf{x}$

$$H(\mathbf{x}) \triangleq -\mathbb{E}\left[\log_2 p(\mathbf{x})\right], \tag{2.3}$$

$H(\mathbf{x},\mathbf{y})$ is the joint entropy rate between channel input $\mathbf{x}$ and channel output $\mathbf{y}$

$$H(\mathbf{x},\mathbf{y}) \triangleq -\mathbb{E}\left[\log_2 p(\mathbf{x},\mathbf{y})\right], \tag{2.4}$$

and $H(\mathbf{y})$ is the entropy rate of the received random discrete-time process $\mathbf{y}$

$$H(\mathbf{y}) \triangleq -\mathbb{E}\left[\log_2 p(\mathbf{y})\right]. \tag{2.5}$$

2.1 SISO scenario

In [19] it is described a method to compute the mutual information of a finite-state hidden Markov model employing the forward recursion of the well-known BCJR algorithm [20]. Such a method can be extended to all channel models with an infinite number of states, like AWGN channels affected by phase noise, finding an auxiliary finite-state channel well approximating the real one; hence the algorithms allows to compute a lower bound of the actual mutual information, which tends to such value when the number of states of the auxiliary channel grows to infinity.

In this section we extend the approach described in [19] to evaluate the mutual information for a SISO channel impaired by phase noise. The channel model is the following

$$y_k = x_k e^{j\theta_k} + w_k, \quad k = 1,2,\ldots,n \tag{2.6}$$

where x_k and y_k are the transmitted and received symbols,respectively, at time instant k, w_k is white Gaussian noise and θ_k is the phase noise sample at instant k, defined as the sum of phase noise at transmitter and at receiver. Phase noise is modeled as a Wiener process, as shown in the Section 1.2.1.

We consider the problem of computing the information rate between the input process $\mathbf{x}^n = (x_1, x_2, \ldots, x_n)$, made by independent and uniformly distributed symbols, and the corresponding output process $\mathbf{y}^n = (y_1, y_2, \ldots, y_n)$ with the channel model described in (4.1). Let $\{\theta\} = (\theta_1, \theta_2, \ldots, \theta_n)$ is the phase process representing the evolution of phase noise, we will assume that the state θ_k takes values in some finite set and that the process $\{\theta\}$ is ergodic; a sufficient condition for ergodicity is that $p(\theta_k|\theta_1) > 0$ for all sufficiently large k.

Recalling the last expression of mutual information in (2.2) the computation of entropy rate $H(\mathbf{x})$ and the differential entropy rate $H(\mathbf{x}|\mathbf{y})$ can be carried out by simulation way thanks to the Shannon-MacMillian-Breimann theorem [14, Theorem 16.8.1] which ensures the convergence with probability one of

$$H(\mathbf{y}) = -\lim_{n\to\infty} \frac{1}{n} \log p(\mathbf{y}^n) \tag{2.7}$$

$$H(\mathbf{x},\mathbf{y}) = -\lim_{n\to\infty} \frac{1}{n} \log p(\mathbf{x}^n,\mathbf{y}^n). \tag{2.8}$$

if $\mathbf{x}^n$ and $\mathbf{y}^n$ are finite-valued stationary ergodic processes. The entropy rate $H(\mathbf{x})$, finally, can be analytically evaluated. By replacing (2.7) and (2.8) in (2.2) we get

$$I(\mathbf{x};\mathbf{y}) \simeq H(x) - \frac{1}{n} \log p(\mathbf{y}^n) + \frac{1}{n} \log p(\mathbf{x}^n,\mathbf{y}^n). \tag{2.9}$$

Hence, from (2.9) it is clear that in order to compute the mutual information is sufficient to obtain the values of probability $p(\mathbf{y}^n)$ and $p(\mathbf{x}^n,\mathbf{y}^n)$. Such values can

be effectively computed by the forward recursion of the BCJR algorithm, employed to implement a maximum a posteriori probability (MAP) symbol detection strategy [20]. In order to have a finite representation of the channel we discretize the values that samples θ_k may assume in N levels. Obviously, we are approximating the real channel and, by increasing N, better approximation can be achieved at the price of higher channel state cardinality. In particular we propose an uniform discretization, by which samples θ_k are considered belonging to the below alphabet

$$\theta_k \in \left\{ i\frac{2\pi}{N} \right\} \quad \forall i = 0, \ldots, N-1 \tag{2.10}$$

The number of discretization phase levels N represents the number of states of the trellis which the algorithm is based on.

Lets consider the computation of $p(\mathbf{y}^n)$ in (2.9). By defining $\mu_{1,k}(\theta_k) = p(\mathbf{y}^k, \theta_k)$ the state metric at time instant k associated to the phase state θ_k, we obtain, taking the average on all possible phase states at time $k-1$ and on all symbols of constellation

$$\begin{aligned}
\mu_{1,k}(\theta_k) &= \sum_{\mathbf{x}_k} \sum_{\theta_{k-1}} p(\mathbf{y}^k, \theta_k, x_k, \theta_{k-1}) \\
&= \sum_{\mathbf{x}_k,} \sum_{\theta_{k-1}} p(y_k|\mathbf{y}^{k-1}, \theta_k, \theta_{k-1}, x_k) p(\mathbf{y}^{k-1}, \theta_k, \theta_{k-1}, x_k) \\
&= \sum_{\mathbf{x}_k,} \sum_{\theta_{k-1}} p(y_k|\theta_k, x_k) p(\theta_k|\theta_{k-1}) p(\mathbf{y}^{k-1}, \theta_{k-1}) p(x_k) \\
&= \sum_{\mathbf{x}_k,} \sum_{\theta_{k-1}} p(y_k|\theta_k, x_k) p(\theta_k|\theta_{k-1}) \mu_{1,k-1}(\theta_{k-1}) p(x_k)
\end{aligned} \tag{2.11}$$

In (2.11) the probability $p(y_k|\theta_k, x_k)$ is the probability of received symbols, defined as

$$p(y_k|\theta_k, x_k) = \frac{1}{\sqrt{2\sigma^2}} \exp\left\{ -\frac{|y_k - x_k e^{j\theta_k}|^2}{2\sigma^2} \right\}, \tag{2.12}$$

$p(\theta_k|\theta_{k-1})$ represents the markovian transition probability of phase process described in Section 1.2.1

$$p(\theta_k|\theta_{k-1}) = \int_{\Delta_\theta} \frac{1}{\sqrt{2\sigma_\theta^2}} \exp\left\{ -\frac{(\theta_k - \theta_{k-1})^2}{2\sigma_\theta^2} \right\} d\theta_k, \tag{2.13}$$

$\mu_{1,k-1}(\theta_{k-1})$ is the state metric at the previous step and $p(x_k)$ is the a-priori probability of transmitted symbols (in the following we will neglect it because the source is defined as uniformly distributed). At final trellis step n we can finally compute

$p(\mathbf{y}^n)$ as the sum of all final state metrics

$$p(\mathbf{y}^n) = \sum_{\theta_n} \mu_{1,n}(\theta_n). \tag{2.14}$$

Moving to logarithm domain we get

$$\begin{aligned}\log \mu_{1,k}(\theta_k) &\propto \log \sum_{\mathbf{x}_k,\ \theta_{k-1}} \sum p(y_k|\theta_k, x_k) p(\theta_k|\theta_{k-1}) \mu_{1,k-1}(\theta_{k-1}) \\ &\propto \log \sum_{\mathbf{x}_k, \theta_{k-1}} \exp\left\{-|y_k - x_k e^{j\theta_k}|^2 + \log p(\theta_k|\theta_{k-1}) + \log \mu_{1,k-1}(\theta_{k-1})\right\}.\end{aligned} \tag{2.15}$$

Defining the max* operator as

$$\begin{aligned}\max_{a_i}{}^* \{a_1, \ldots, a_N\} &= \log \sum_{i=1}^{N} \exp(a_i) \\ &= \max_i \{a_i\} + \log\left\{1 + \exp\left(-\left|\max_i \{a_i\} - \min_i \{a_i\}\right|\right)\right\}\end{aligned} \tag{2.16}$$

from (2.15) and (2.14) we can derive the expression

$$\log \mu_{1,k}(\theta_k) \propto \max_{\mathbf{x}_k, \theta_{k-1}}{}^* \left\{-|y_k - x_k e^{j\theta_k}|^2 + \log p(\theta_k|\theta_{k-1}) + \log \mu_{1,k-1}(\theta_{k-1}\right\} \tag{2.17}$$

$$\log p(\mathbf{y}^n) = \max_{\theta_n}{}^* \log \mu_{1,\theta_n}(\theta_n). \tag{2.18}$$

The computation of $p(\mathbf{y}^n, \mathbf{x}^n)$ in (2.9) is similar to the previous derivation. By defining $\mu_{2,k}(\theta_k) = p(\mathbf{y}^k, \mathbf{x}^k, \theta_k)$ a second state metric at instant k, taking only the average on all possible phase states at instant $k-1$, we obtain

$$\begin{aligned}\mu_{2,k}(\theta_k) &= \sum_{\theta_{k-1}} p(\mathbf{y}^k, \mathbf{x}^k, \theta_k, \theta_{k-1}) \\ &= \sum_{\theta_{k-1}} p(y_k|\mathbf{y}^{k-1}, \theta_k, \theta_{k-1}, \mathbf{x}^k) p(\mathbf{y}^{k-1}, \theta_k, \theta_{k-1}, \mathbf{x}^k) \\ &= \sum_{\theta_{k-1}} p(y_k|\theta_k, x_k) p(\theta_k|\theta_{k-1}) p(\mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \theta_{k-1}) p(x_k) \\ &= \sum_{\theta_{k-1}} p(y_k|\theta_k, x_k) p(\theta_k|\theta_{k-1}) \mu_{2,k-1}(\theta_{k-1}) p(x_k)\end{aligned} \tag{2.19}$$

and at final trellis step n we finally have

$$p(\mathbf{y}^n, \mathbf{x}^n) = \sum_{\theta_n} \mu_{2,n}(\theta_n). \tag{2.20}$$

Moving to the logarithmic domain and using the definition (2.16) in (2.19) and in (2.20) we obtain the expressions

$$\log \mu_{2,k}(\theta_k) \propto \max_{\theta_{k-1}}{}^{*} \left\{ -|y_k - x_k e^{j\theta_k}|^2 + \log p(\theta_k|\theta_{k-1}) + \log \mu_{2,k-1}(\theta_{k-1}) \right\} \tag{2.21}$$

$$\log p(\mathbf{x}^n, \mathbf{y}^n) = \max_{\theta_n}{}^{*} \log \mu_{2,\theta_n}(\theta_n). \tag{2.22}$$

2.1.1 Simulation results

In this Section we show simulation results for the channel model described in Section 2.1. For simulations we used $N = 300$ quantization levels of phase noise and sequences of $n = 5000$ symbols. Phase noise is modeled as a Wiener process with $\sigma_\Delta = 6°$.

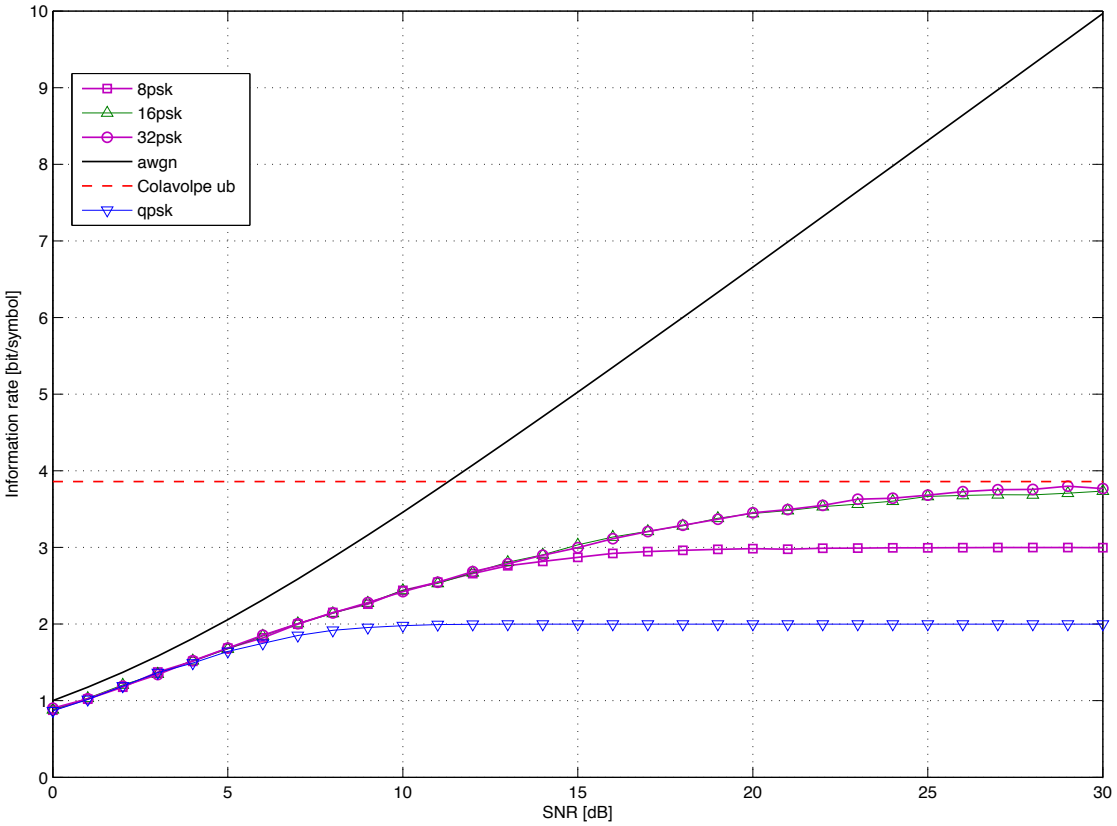

Figure 2.1. Lower bound to capacity for PSK modulation with several cardinalities (4-PSK, 8-PSK, 16-PSK) related to a Wiener phase noise model with $\sigma_\Delta = 6°$, AWGN capacity, the upper bound to capacity for M-PSK modulations computed in [1]

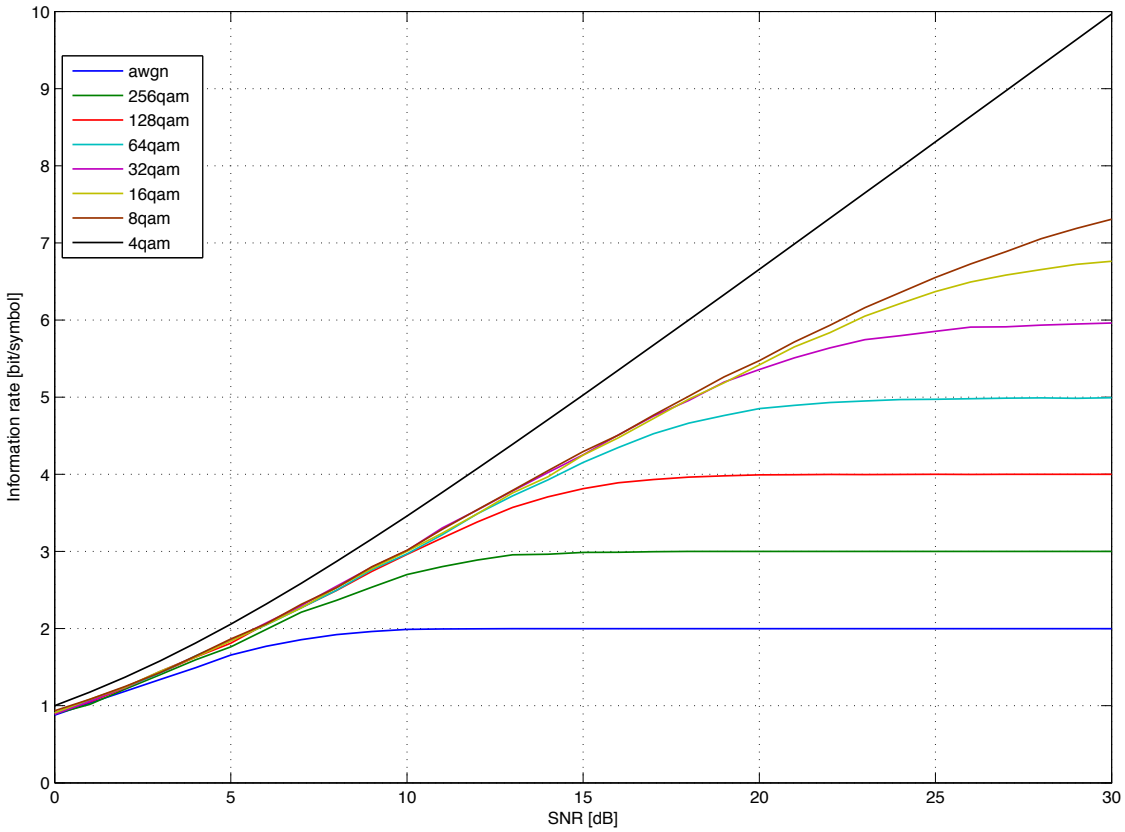

Figure 2.2. Lower bound to capacity for QAM modulation with several cardinalities (4-QAM, 8-QAM, 16-QAM, 32-QAM, 64-QAM, 128-QAM, 256-QAM related to a Wiener phase noise model with $\sigma_\Delta = 6°$ and AWGN capacity

In Figure 2.1 is depicted the mutual information for M-PSK modulations. As you can see, phase noise has no effect on 4-PSK and 8-PSK modulations at high SNR values, but for higher-order modulations, like 16-PSK, the information rate is limited at about 3.8 bit/symbol. This result is perfectly in accordance with the conclusions in [1], in which Barbieri *et al.* show a theoretical asymptotic upper bound to the information rate at high SNR for M-PSK at 3.85 bit/symbol.

In Figure 2.2 we show the mutual information of a channel carrying symbols belonging to M-QAM modulation. As shown also in [1] QAM modulations outperform the PSKs and are not saturated. QAM modulations seems to be less sensitive to the time-varying phase noise, because the information is conveyed by the amplitude and in the following we will use those results related to QAM constellations as a lower bound to channel capacity.

2.2 MIMO scenario

We now extend the approach described in Section 2.1 to the 2×2 MIMO LoS system impaired by phase noise described in 1.3. The input-output relation at time instant k is the following

$$\mathbf{y}_k = e^{j\theta_k}\mathbf{H}\mathbf{x}_k + \mathbf{w}_k. \tag{2.23}$$

We consider now the problem of computing the information rate between the $2n$-length input sequence $\mathbf{x}^n = (\mathbf{x}_1^\intercal, \mathbf{x}_2^\intercal, \ldots, \mathbf{x}_n^\intercal)$, whose components are 2-dimensional input vectors of independent and uniformly distributed symbols, and the corresponding output sequence $\mathbf{y}^n = (\mathbf{y}_1^\intercal, \mathbf{y}_2^\intercal, \ldots, \mathbf{y}_n^\intercal)$.

Phase noise is modeled as a Wiener process (see Section 1.2.1) and for this channel we need again to evaluate the probability of received sequence $p(\mathbf{y}^n)$ and the joint probability $p(\mathbf{x}^n, \mathbf{y}^n)$ in (2.2). In particular, by defining $\tilde{\mu}_{1,k}(\theta_k) = p(\mathbf{y}^k, \theta_k)$ and $\tilde{\mu}_{2,k}(\theta_k) = p(\mathbf{x}^n, \mathbf{y}^k, \theta_k)$ we obtain

$$\begin{aligned}
\tilde{\mu}_{1,k}(\theta_k) &= \sum_{\mathbf{x}_k}\sum_{\theta_{k-1}} p(\mathbf{y}^k, \theta_k, \mathbf{x}_k, \theta_{k-1}) \\
&= \sum_{\mathbf{x}_k}\sum_{\theta_{k-1}} p(\mathbf{y}_k|\theta_k, \mathbf{x}_k)p(\theta_k|\theta_{k-1})\mu_{1,k-1}(\theta_{k-1})p(\mathbf{x}_k), \qquad (2.24)\\
\tilde{\mu}_{2,k}(\theta_k) &= \sum_{\theta_{k-1}} p(\mathbf{y}^k, \mathbf{x}^k, \theta_k, \theta_{k-1}) \\
&= \sum_{\theta_{k-1}} p(\mathbf{y}_k|\theta_k, \mathbf{x}_k)p(\theta_k|\theta_{k-1})\mu_{2,k-1}(\theta_{k-1})p(\mathbf{x}_k). \qquad (2.25)
\end{aligned}$$

Moving to the logarithm domain we can compute the entropy rate of received sequence $H(\mathbf{y}^n)$ and the joint entropy rate $H(\mathbf{x}^n, \mathbf{y}^n)$ in (2.2), to evaluate the mutual information, in the following way

$$H(\mathbf{y}^n) = -\frac{1}{n}\max_{\theta_n}{}^{*}\, \log \tilde{\mu}_{1,\theta_n}(\theta_n), \tag{2.26}$$

$$H(\mathbf{x}^n, \mathbf{y}^n) = -\frac{1}{n}\max_{\theta_n}{}^{*}\, \log \tilde{\mu}_{2,\theta_n}(\theta_n) \tag{2.27}$$

where $\tilde{\mu}_{1,\theta_n}(\theta_n)$ and $\tilde{\mu}_{2,\theta_n}(\theta_n)$ are the state metrics at final time interval n.

2.2.1 Simulation results

In Figure 2.3 we present the mutual information rate of the described 2×2 MIMO LoS channel for 16-QAM, 64-QAM and 256-QAM constallations. The considered channel is the one we showed in (2.23) with unitary channel matrix (1.12). Oscillators are shared among the antennas as we discussed in Section 1.3 and standard deviation of Gaussian increment of Wiener process is $\sigma_\Delta = 6^\circ$.

For MIMO channel, as well as the SISO scenario, the mutual information of QAM modulations does not saturate and can be a reasonable lower bound to channel capacity. At low SNR region the upper bound to capacity is tight to the AWGN capacity and the loss increases in high SNR region.

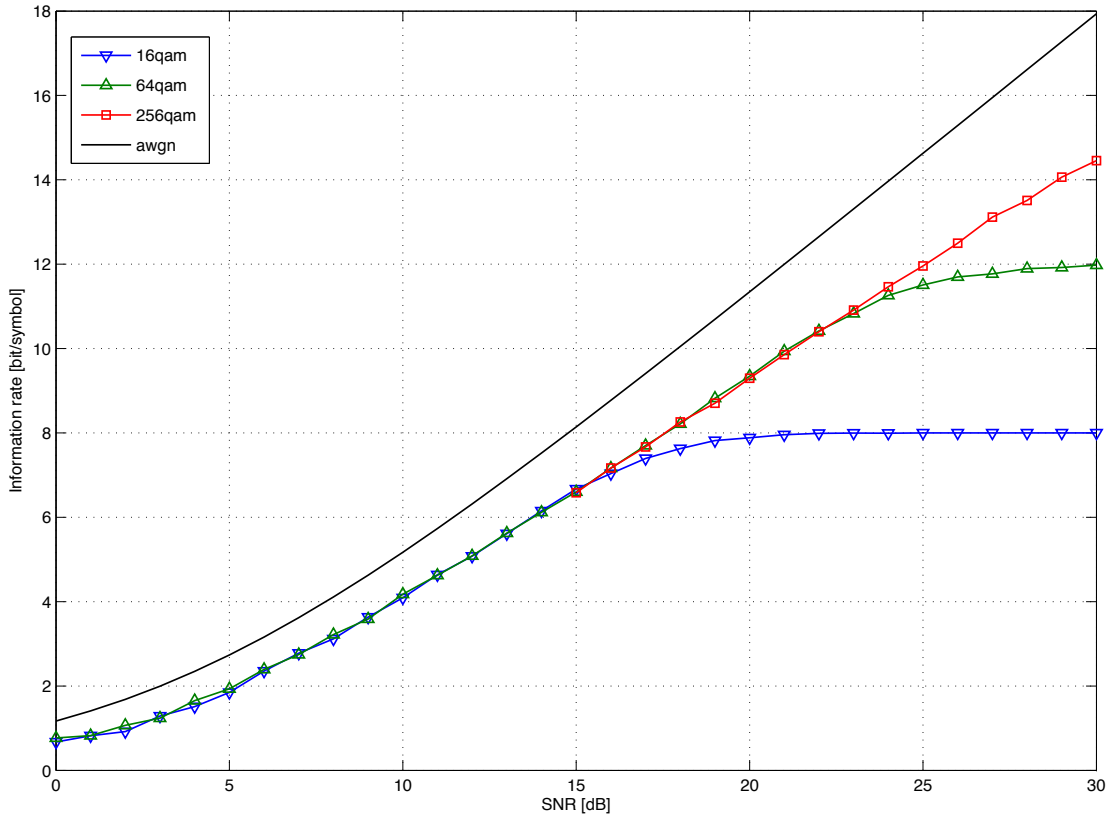

Figure 2.3. Lower bound to capacity for 16-QAM, 64-QAM and 256-QAM related to a 2x2 MIMO channel with shared oscillators (phase noise is modeled with a Wiener process with $\sigma_\Delta = 6°$) and AWGN capacity

Chapter 3

Upper bounds to capacity

In this Chapter we present a semi-analytic expression of non-asymptotic capacity upper bound for a multiple-antenna systems (we treat the SISO scenario as a particular case of MIMO channel) impaired by a phase noise modeled as a Wiener process (cfr. Section 1.2.1). We consider, again, the case where a single oscillator is shared among the antennas at each transceiver.

In Section 3.1 we study the case with average energy constraint, while in Section 3.2 we consider a more stringent constraint on peak-power.

3.1 Upper bound with average energy constraint

Lapidoth [21] obtained a large-SNR characterization of the capacity of the general class of stationary phase-noise channels (as we stated in 1.2.1 the widely used Wiener model belongs to this class). Specifically, he showed that the capacity of the phase-noise channel is asymptotically equal to half the capacity of an AWGN channel with the same SNR plus a correction term that accounts for the memory in the phase-noise process. The result in [21] has been recently extended to the *waveform* phase-noise channel in [13]. The capacity of the block-memoryless phase-noise channel (a non-stationary channel) has been characterized in [22] in the large-SNR regime.

Moving away from asymptotic results, Katz and Shamai [23] provided tight upper and lower bounds on the capacity of memoryless phase noise channels. These bounds have been recently extended to the block-memoryless phase-noise case in [24]. For the Wiener phase-noise model, an upper bound on the rates achievable with PSK constellations has been recently proposed in [1].

The impact of phase noise on multiple-antenna systems has been recently discussed in [3] where it is shown that different RF circuitries configurations (e.g., independent oscillators at each antenna as opposed to a single oscillator driving all antennas) yield different capacity behavior at high SNR.

In this Section we study the capacity of multiple-antenna systems affected by phase noise [2]. Specifically, we consider the scenario where a single oscillator drives all RF circuitries at each transceiver presenting a non-asymptotic capacity upper bound for the case of Wiener phase noise. When particularized to constant modulus constellations and to single-antenna systems, our bound recovers the upper bound obtained in [1]. By exhibiting a matching lower bound, we show that our upper bound is tight in the large-SNR regime.

Focusing on single-antenna systems, we finally compare our upper bound with lower bounds obtained by evaluating numerically the information rates achievable with QAM constellations and discussed in Chapter 2. For the case of a Wiener phase-noise process with standard deviation of the phase increments equal to 6°, our results imply that QAM constellations incur a penalty of more than 3 dB for medium/high SNR values.

We consider now a generic $M \times M$ MIMO phase noise channel with memory, described by the following input-output relation

$$\mathbf{y}_k = e^{j\theta_k}\mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad k = 1,2,\dots,n \tag{3.1}$$

Here, $\mathbf{x}_k$ denotes the M-dimensional input vector at discrete time k, $\mathbf{H}$ is the MIMO channel matrix, which we assume deterministic, full-rank, and known to the transmitter and the receiver, $\{\theta_k\}$ is the phase-noise process modeled as a Wiener model, and $\mathbf{w}_k$ is the additive Gaussian noise, which is circularly symmetric with zero mean and covariance matrix $\mathbf{I}_M$, i.e., $\mathbf{w}_k \sim \mathcal{NC}(\mathbf{0}, \mathbf{I}_M)$. This model is accurate for MIMO LoS systems (see Section 1.3).

We are interested in computing the capacity of the MIMO phase-noise channel (3.1), which is defined as

$$C(\rho) = \lim_{n\to\infty} \frac{1}{n} \sup I(\mathbf{x}^n; \mathbf{y}^n). \tag{3.2}$$

Here, the supremum is over all probability distributions on the transmitted sequence $\mathbf{x}^n = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ that satisfy the average-power constraint

$$\sum_{k=1}^{n} \mathbb{E}||\mathbf{x}_k||^2 \leq n\rho. \tag{3.3}$$

Since the additive noise has unit variance, the parameter $\rho \geq 0$ can be thought of as the SNR. The capacity $C(\rho)$ is not known in closed form. In the following we shall present a capacity upper bound that will turn out to be tight in the large-SNR regime.

Before presenting our upper bound, two observations are in order.

i) As $\mathbf{H}$ is known to transmitter and receiver, $C(\rho)$ depends on $\mathbf{H}$ only through its singular values. For simplicity, in the remainder of the paper we shall focus on the special case when all eigenvalues of $\mathbf{H}$ are equal to one. In this case, we can (and will) assume without loss of generality $\mathbf{H} = \mathbf{I}_M$.

ii) The following proposition establishes that the capacity-achieving input process $\{\mathbf{x}_k\}$ can be assumed isotropically distributed, a property that will be useful in our analysis.

Proposition 1. *The input process $\{\mathbf{x}_k\}$ that achieves the capacity of the channel (3.1), with $\mathbf{H} = \mathbf{I}_M$, can be assumed isotropically distributed. Specifically, if $\{\mathbf{x}_k\}$ achieves $C(\rho)$ in (3.2) then $\{\mathbf{U}_k\mathbf{x}_k\}$, where the matrix-valued random process $\{\mathbf{U}_k\}$ is iid and each $\mathbf{U}_k$ is uniformly distributed on the set of $M \times M$ unitary matrices, achieves $C(\rho)$ as well.*

Proof. The proof, which exploits that $\mathbf{U}_k\mathbf{w}_k \sim \mathbf{w}_k$, follows the same steps as the proof of [25, Prop. 7]. □

We next present an upper bound on $C(\rho)$, which is constructed by extending to the MIMO case the method used in [21] to derive an asymptotic bound on the capacity of stationary single-antenna phase-noise channels. We also use the approach proposed in [24] to make the bound non-asymptotic.

Theorem 1. *The capacity of the channel (3.1) can be upper-bounded a $C(\rho) \leq U(\rho)$, where*

$$U(\rho) \triangleq \min_{\alpha>0} \min_{\lambda\geq 0} \{\alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) + \max_{\xi\geq 0} g_{\lambda,\alpha}(\xi, \rho)\}. \tag{3.4}$$

Here,

$$\begin{aligned} g_{\lambda,\alpha}(\xi,\rho) \triangleq &(M-\alpha)\mathbb{E}\left[\log\left(|\xi + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\ &- h(|\xi + z|^2) + (\alpha - \lambda)\frac{\xi^2 + M}{\rho + M} \\ &- h\left(\underline{/\xi + z} + \Delta||\xi + z|\right) \end{aligned} \tag{3.5}$$

where $\underline{/x}$ denotes the phase of $x \in \mathbb{C}$ and $z, z_1, \ldots, z_M$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables. Furthermore,

$$d_{\lambda,\alpha} \triangleq \log \frac{\Gamma(\alpha)}{\Gamma(M)} + \lambda - M + 1. \tag{3.6}$$

Proof. Because of Proposition 1, we can restrict ourselves to input processes that are isotropically distributed. Specifically, we will consider $\{\mathbf{x}_k\}$ of the form $\{\mathbf{x}k = s_k\mathbf{v}_k\}$, where $s_k = ||\mathbf{x}_k||$ and $\mathbf{v}_k = \mathbf{x}_k/s_k$, with $\mathbf{v}_k$ uniformly distributed on the unit sphere in $\mathbb{C}^M$ and independent of s_k.

We start by using chain rule as follows

$$I(\mathbf{x}^n;\mathbf{y}^n) = \sum_{k=1}^{n} I(\mathbf{x}^n;\mathbf{y}_k|\mathbf{y}^{k-1}). \tag{3.7}$$

By proceeding as in [21], we next upper-bound each term on the right-hand side (RHS) of (3.7) as

$$\begin{aligned}
I(\mathbf{x}^n;\mathbf{y}_k|\mathbf{y}^{k-1}) &= h(\mathbf{y}_k|\mathbf{y}^{k-1}) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^n)\\
&\overset{(a)}{\leq} h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^n)\\
&= h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k)\\
&\overset{(b)}{\leq} h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k,\theta^{k-1})\\
&\overset{(c)}{=} h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{x}_k,\theta^{k-1})\\
&= h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{x}_k) + h(\mathbf{y}_k|\mathbf{x}_k)\\
&\quad - h(\mathbf{y}_k|\mathbf{x}_k,\theta^{k-1})\\
&= I(\mathbf{x}_k;\mathbf{y}_k) + I(\mathbf{y}_k;\theta^{k-1}|\mathbf{x}_k)\\
&\overset{(d)}{=} I(\mathbf{x}_k;\mathbf{y}_k) + I(\mathbf{y}_k;\theta_{k-1}|\mathbf{x}_k).
\end{aligned} \tag{3.8}$$

Here, in (a) and (b) we used that conditioning reduces entropy; (c) follows because $\mathbf{y}_k$ and the pair $(\mathbf{y}^{k-1},\mathbf{x}^{k-1})$ are conditionally independent given $(\theta^{k-1},\mathbf{x}_k)$; finally, (d) holds because $\{\theta_k\}$ is a first-order Markov process. Let $z_k \sim \mathcal{CN}(0,1)$. The second term on the RHS of (3.8) can be evaluated as follows:

$$\begin{aligned}
I(\mathbf{y}_k;\theta_{k-1}|\mathbf{x}_k) &\overset{(a)}{=} I(e^{j\theta_k}s_k + z_k;\theta_{k-1}|s_k)\\
&\overset{(b)}{=} I\big(e^{j\theta_k}(s_k+z_k);\theta_{k-1}|s_k\big)\\
&= I\big(|s_k+z_k|,\underline{\angle s_k+z_k} + \theta_k;\theta_{k-1}|s_k\big)\\
&\overset{(c)}{=} I(\underline{\angle s_k+z_k} + \theta_k;\theta_{k-1}||s_k+z_k|,s_k)\\
&= h(\underline{\angle s_k+z_k} + \theta_k||s_k+z_k|,s_k)\\
&\quad - h(\underline{\angle s_k+z_k} + \theta_k||s_k+z_k|,\theta_{k-1},s_k)\\
&\overset{(d)}{=} \log(2\pi)\\
&\quad - h(\underline{\angle s_k+z_k} + \Delta||s_k+z_k|,s_k).
\end{aligned} \tag{3.9}$$

Here, (a) follows because $\mathbf{v}_k^\dagger \mathbf{y}_k \sim e^{j\theta_k} s_k + z_k$ is a sufficient statistics for θ_{k-1}; (b) follows because z_k is circularly symmetric; (c) holds because $|s_k + z_k|$ and θ_{k-1} are independent; finally, (d) holds because $\theta_k \sim \mathcal{U}[0,2\pi]$ and because of (1.1).

Substituting (3.9) into (3.8), then (3.8) into (3.7), and using that $\{\theta_k\}$ is a stationary process, we get

$$C(\rho) \leq \sup_{\mathcal{Q}_s} \{I(s, \mathbf{v}; \mathbf{y}) + \log(2\pi) - h(\underline{\angle s+z} + \Delta || s+z|, s)\} \tag{3.10}$$

where

$$\mathbf{y} = e^{j\theta} s\mathbf{v} + \mathbf{w} \tag{3.11}$$

with $\theta \sim \mathcal{U}[0,2\pi]$, $\mathbf{v}$ uniformly distributed on the unit sphere in $\mathbb{C}^M$, $z \sim \mathcal{CN}(0,1)$, and Δ distributed as in (1.3); the supremum in (3.10) is over all distributions $\mathcal{Q}_s$ on $s \geq 0$ that satisfy $\mathbb{E}[s^2] \leq \rho$.

We further upper-bound the first term on the RHS of (3.10) (which corresponds to the mutual information achievable on a memoryless channel with uniform phase noise) by using duality [26, Thm. 5.1] and obtain that for every $\mathcal{Q}_s$ and for every $\alpha > 0$ and $\lambda > 0$ (see Appendix A)

$$\begin{aligned} I(s, \mathbf{v}; \mathbf{y}) \leq{}& \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} \\ &+ (M-\alpha)\mathbb{E}\left[\log\left(|s+z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\ &- h(|s+z|^2|s) + (\alpha - \lambda)\frac{\mathbb{E}\left[s^2\right] + M}{\rho + M}. \end{aligned} \tag{3.12}$$

Here, $d_{\lambda,\alpha}$ is the constant defined in (3.6) and $z, z_1, \ldots, z_M$ are i.i.d. $\mathcal{U}[0,2\pi]$-distributed random variables. Substituting (3.12) into (3.10), we obtain

$$\begin{aligned} C(\rho) &\leq \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) \\ &\quad + \sup_{\mathcal{Q}_s} \Bigg\{ (M-\alpha)\mathbb{E}\left[\log\left(|s+z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\ &\quad - h(|s+z|^2|s) + (\alpha - \lambda)\frac{\mathbb{E}[s^2] + M}{\rho + M} \\ &\quad - h(\underline{\angle s+z} + \Delta || s+z|, s) \Bigg\} \\ &\leq \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) + \max_{\xi \geq 0} g_{\lambda,\alpha}(\xi, \rho). \end{aligned} \tag{3.13}$$

where $g_{\lambda,\alpha}(\xi,\rho)$ was defined in (3.5). In the last step, we upper-bounded the supremum over $\mathcal{Q}_s$ with the supremum over all deterministic $\xi \geq 0$. The resulting upper bound can be tightened by minimizing it over α and λ, which yields (3.4).

This concludes the proof. □

It is instructive to note that if we further lower-bound the last term on the RHS of (3.10) as

$$h(\underline{\angle s+z}+\Delta||s+z|,s) \geq h(\underline{\angle s+z}+\Delta||s+z|,s,\underline{\angle s+z}) = h(\Delta) \tag{3.14}$$

we obtain

$$C(\rho) \leq \sup_{\mathcal{Q}_s}\big\{I(s,\mathbf{v};\mathbf{y})\big\} + \log(2\pi) - h(\Delta) \tag{3.15}$$

where $s\mathbf{v}$ is the input to the memoryless phase-noise channel with uniform phase noise (3.11). The inequality (3.15) can be interpreted as follows: the capacity of a Wiener phase-noise channel is upper-bounded by the capacity of a memoryless phase-noise channel with uniform phase noise, plus a correction term that accounts for the memory in the channel and does not depend on the SNR ρ. If we now specialize (3.15) to single antenna systems and we add the additional constraint on $\mathcal{Q}_s$ that $|s|^2 = \rho$ with probability one (which holds, for example, if a PSK constellation is used), the first term on the RHS of (3.15) vanishes and we recover the upper bound previously reported in [1, Theorem 2].

3.1.1 Large-SNR Regime

In Theorem 4 below, we present an asymptotic characterization of $C(\rho)$, which generalizes to the MIMO case the asymptotic characterization reported in [21] for the single-antenna case.

Theorem 2. *In the large-SNR regime, the capacity of the Wiener phase-noise channel* (3.1) *behaves as*

$$C(\rho) = \left(M-\frac{1}{2}\right)\log\frac{2\rho}{2M-1} + \log\frac{\Gamma(M-1/2)}{\Gamma(M)} + \frac{1}{2}\log\pi - h(\Delta) + o(1) \tag{3.16}$$

where $o(1)$ *indicates a function of* ρ *that vanishes in the limit* $\rho \to \infty$.

Proof. The asymptotic characterization (3.16) is obtained by proving that an appropriately modified version of the upper bound previously presented matches the lower bound we shall report in this section up to a $o(1)$ term.

Upper bound We exploit the property that the high-SNR behavior of $C(\rho)$ does not change if the support of the input distribution is constrained to lie outside a sphere of arbitrary radius. This result, known as escape-to-infinity property of the capacity-achieving input distribution [26, Def. 4.11], is formalized in the following lemma.

Lemma 1. *Fix an arbitrary $\xi_0 > 0$ and let $\mathcal{K}(\xi_0) = \{\mathbf{x} \in \mathcal{C}^M : \|\mathbf{x}\| \geq \xi_0\}$. Denote by $C^{(\xi_0)}(\rho)$ the capacity of the channel (3.1) when the input signal is subject to the average-power constraint (3.3) and to the additional constraint that $\mathbf{x}_k \in \mathcal{K}(\xi_0)$ almost surely for all k. Then*

$$C(\rho) = C^{(\xi_0)}(\rho) + o(1), \quad \rho \to \infty \tag{3.17}$$

with $C(\rho)$ given in (3.2).

Proof. The lemma follows directly from [25, Thm. 8] and [26, Thm. 4.12]. □

Fix $\xi_0 > 0$. By performing the same steps leading to (3.13), but accounting for the additional constraint that $\mathbf{x} \in \mathcal{K}(\xi_0)$ almost surely and also setting $\alpha = \lambda = M - 1/2$, we obtain: $C^{(\xi_0)}(\rho) \leq U^{(\xi_0)}(\rho)$, where

$$\begin{aligned} U^{(\xi_0)}(\rho) \triangleq \left(M - \frac{1}{2}\right) \log \frac{2(\rho + M)}{2M - 1} + \log \frac{\Gamma(M - 1/2)}{\Gamma(M)} \\ + \log(2\pi) + \frac{1}{2} + \max_{\xi \geq \xi_0} \{\tilde{g}(\xi)\}. \end{aligned} \tag{3.18}$$

with $\tilde{g}(\xi) \triangleq g_{\lambda,\alpha}(\xi, \rho)|\lambda = \alpha = M - 1/2$. As $\lim_{\xi \to \infty} \tilde{g}(\xi) = -(1/2)\log(4\pi e) - h(\Delta)$ (see [21, Eq. (9)] and proceed similarly to the proof of [26, Lemma 6.9]), we can make (3.18) to be arbitrarily close to (3.16) by choosing ξ_0 sufficiently large.

Lower bound Take $\{\mathbf{x}_k\}$ i.i.d. and isotropically distributed with

$$\|\mathbf{x}_k\|^2 = \rho \frac{t_{\rho,\rho_0}}{(M - 1/2)} \tag{3.19}$$

where, for a given $\rho_0 > 0$, the random variable t_{ρ,ρ_0} has pdf

$$f_{t_{\rho,\rho_0}}(a) = \begin{cases} \dfrac{f^{(\rho,\rho_0)}(a)}{\Pr\{z > \rho_0/\rho\}}, & \text{if } a > \rho_0/\rho \\ 0, & \text{otherwise.} \end{cases} \tag{3.20}$$

Here, $f^{(\rho,\rho_0)}$ denotes the pdf of a random variable that follows a $\Gamma((M - 1/2) \cdot \Pr\{z > \rho_0/\rho\},1)$ distribution. Let $t \sim \Gamma(M - 1/2,1)$ and denote its pdf by f_t. Note that for all ρ_0 the pdf $f_{t_{\rho,\rho_0}}$ converges point-wise to f_t as $\rho \to \infty$. As

$$\mathbb{E}\left[t_{\rho,\rho_0} \leq M - 1/2\right] \tag{3.21}$$

the average-power constraint (3.3) is satisfied. A key feature of the distribution of $\|\mathbf{x}_k\|$ in (3.19) is that $\Pr\{\|\mathbf{x}_k\| < \xi_0\} = 0$ where

$$\xi_0^2 \triangleq \frac{\rho_0}{M - 1/2}. \tag{3.22}$$

Note that $\xi_0 \to \infty$ as $\rho_0 \to \infty$, a property that will be useful in the remainder of the proof. To obtain the desired lower bound, we use chain rule for mutual information and that mutual information is non negative

$$\begin{aligned} I(\mathbf{x}^n; \mathbf{y}^n) &= \sum_{k=1}^{n} I(\mathbf{x}_k; \mathbf{y}^n | \mathbf{x}^{k-1}) \\ &\geq \sum_{k=2}^{n} I(\mathbf{x}_k; \mathbf{y}^k | \mathbf{x}^{k-1}). \end{aligned} \tag{3.23}$$

Fix now $k \geq 2$ and set

$$\epsilon_k \triangleq I(\mathbf{x}_k; \theta_{k-1} | \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}). \tag{3.24}$$

We have

$$\begin{aligned} I(\mathbf{x}_k; \mathbf{y}^k | \mathbf{x}^{k-1}) &\overset{(a)}{=} I(\mathbf{x}_k; \mathbf{y}^k, \mathbf{x}^{k-1}) \\ &\overset{(b)}{\geq} I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}) \\ &= I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}, \theta_{k-1}) - \epsilon_k \\ &\overset{(c)}{=} I(\mathbf{x}_k; \mathbf{y}_k, \theta_{k-1}) - \epsilon_k \\ &\overset{(d)}{=} I(\mathbf{x}_k; \mathbf{y}_k | \theta_{k-1}) - \epsilon_k \\ &\overset{(e)}{=} I(\mathbf{x}_2; \mathbf{y}_2 | \theta_1) - \epsilon_2. \end{aligned} \tag{3.25}$$

Here, (a) follows because $\{\mathbf{x}_k\}$ are independent; in (b) we used chain rule for mutual information and that mutual information is nonnegative; (c) follows because $\mathbf{x}_k$ and the pair $(\mathbf{y}_{k-1}, \mathbf{x}_{k-1})$ are conditionally independent given $(\theta_{k-1}, \mathbf{y}_k)$; (d) holds because $\mathbf{x}_k$ and θ_{k-1} are independent; finally (e) follows from stationarity. Substituting (3.25) into (3.23), we obtain

$$C(\rho) \geq I(\mathbf{x}_2; \mathbf{y}_2 | \theta_1) - \epsilon_2. \tag{3.26}$$

We next investigate the two terms on the RHS of (3.26) separately. Specifically, we shall show that the first term has the desired asymptotic expansion, while the second term can be made arbitrarily close to zero by choosing ρ_0 in (3.19) sufficiently large.

The first term on the RHS of (3.26) We write

$$I(\mathbf{x}_2;\mathbf{y}_2|\theta_1) = h(\mathbf{y}_2|\theta_1) - h(\mathbf{y}_2|\mathbf{x}_2,\theta_1) \tag{3.27}$$

and bound the two terms separately. For the first term, we have that

$$\begin{aligned} h(\mathbf{y}_2|\theta_1) &\geq h(\mathbf{y}_2|\mathbf{w}_2,\theta_1) \\ &= h(e^{j\theta_2}\mathbf{x}_2|\theta_1) \\ &\overset{(a)}{=} h(\mathbf{x}_2) \\ &\overset{(b)}{=} h(\|\mathbf{x}_2\|^2) + \log\frac{\pi^M}{\Gamma(M)} + (M-1)\mathbb{E}\left[\log\|\mathbf{x}_2^2\|\right] \\ &\overset{(c)}{=} M\log\frac{\rho}{M-1/2} + \log\frac{\pi^M}{\Gamma(M)} \\ &+ h(t_{\rho,\rho_0}) + (M-1)\mathbb{E}\left[\log t_{\rho,\rho_0}\right]. \end{aligned} \tag{3.28}$$

Here, (a) follows because $\mathbf{x}_2$ is isotropically distributed, and, hence, $e^{j\theta_2}\mathbf{x}_2 \sim \mathbf{x}_2$; in (b) we computed the differential entropy in polar coordinates [26, Lemma 6.15 and 6.17];finally, (c) follows from(3.19). For the second term on the RHS of (3.27), we proceed as follows. Let $\mathbf{x}_2 = s_2\mathbf{v}_2$, with $s_2 = \|\mathbf{x}_2\|$ and, hence, $s_2^2 \sim \rho\, t_{\rho,\rho_0}/(M-1/2)$. Furthermore, let $z_2 \sim \mathcal{N}(0,1)$. Then

$$h(\mathbf{y}_2|\mathbf{x}_2,\theta_1) = h(\mathbf{y}_2|s_2,\mathbf{v}_2,\theta_1) = h(e^{j\theta_2}s_2 + z_2|s_2,\theta_1) + \log(\pi e)^{M-1}. \tag{3.29}$$

Now note that

$$\begin{aligned} h(e^{j\theta_2}s_2 + z_2|s_2,\theta_1) &= h(e^{j\theta_2}(s_2+z_2)|s_2,\theta_1) \\ &\overset{(a)}{=} h(e^{j\Delta}(s_2+z_2)|s_2) \\ &\overset{(b)}{=} h(|s_2+z_2|^2|alts_2) \\ &+ h(\underline{\angle s_2+z_2} + \Delta||s_2+z_2|, s_2) - \log 2 \\ &\overset{(c)}{\leq} \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1+\frac{4\rho}{2M-1}t_{\rho,\rho_0}\right]\right)\right] \\ &+ h(\Delta + \underline{\angle s_2+z_2}|s_2) - \log 2. \end{aligned} \tag{3.30}$$

Here, in (a) we used (1.1) and denoted by Δ a random variable distributed as in (1.3); in (b) we evaluated the differential entropy in polar coordinates [26, Lemma 6.15 and 6.16]. Finally, (c) follows because the Gaussian distribution maximizes differential entropy under a variance constraint and because conditioning reduces entropy. Note finally that

$$h(\underline{\angle s_2+z_2} + \Delta|s_2) \leq \max_{\xi\geq\xi_0} h(\underline{\angle \xi+z_2} + \Delta) h(\underline{\angle \xi_0+z_2} + \Delta). \tag{3.31}$$

This term can be made arbitrarily close to $h(\Delta)$ by choosing ρ_0 in (3.22) sufficiently large.

Summarizing, we have shown that

$$\begin{aligned} I(\mathbf{x}_2;\mathbf{y}_2|\theta_1) \geq\; & M\log\frac{2\rho}{2M-1}+\log\frac{\pi^M}{\Gamma(M)}+h(t_{\rho,\rho_0}) \\ & +(M-1)\mathbb{E}\left[\log t_{\rho,\rho_0}\right] \\ & -\frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1+\frac{4\rho}{2M-1}t_{\rho,\rho_0}\right]\right)\right] \\ & -h(\underline{/\xi_0+z_2}+\Delta)-\log\left[2(\pi e)^{M-1}\right] \\ & \overset{(a)}{=}\left(M-\frac{1}{2}\right)\log\frac{2\rho}{2M-1}+\log\frac{\Gamma(M-1/2)}{\Gamma(M)} \\ & +\frac{1}{2}\log\pi-h(\underline{/\xi_0+z_2}+\Delta)+o(1). \end{aligned} \tag{3.32}$$

Here, (a) follows because

$$\begin{aligned} h(t_{\rho,\rho_0}) &= h(t)+o(1) \\ \mathbb{E}\left[\log(t_{\rho,\rho_0})\right] &= \mathbb{E}\left[\log t\right]+o(1) \\ \mathbb{E}\left[\log(1+c\rho t_{\rho,\rho_0})\right] &= \log(c\rho)+\mathbb{E}\left[\log t\right]+\emptyset(1),\forall c>0 \end{aligned}$$

where $t\sim\Gamma(M-1/2,1)$ and because

$$\begin{aligned} &\mathbb{E}\left[\log t\right]=\psi(M-1/2) \\ &h(t)=(3/2-M)\psi(M-1/2)+M-1/2+\log\Gamma(M-1/2) \end{aligned}$$

with $\psi(\cdot)$ denoting Euler's digamma function.

The second term on the RHS of (3.26) Proceeding similarly as in [26, App. IX], we obtain

$$\begin{aligned} I(\mathbf{x}_2;\theta_1|\mathbf{y}_2,\mathbf{y}_1,\mathbf{x}_1) &= h(\theta_1|\mathbf{y}_2,\mathbf{y}_1,\mathbf{x}_1)-h(\theta_1|\mathbf{y}_2,\mathbf{x}_2,\mathbf{y}_1,\mathbf{x}_1) \\ &\leq h(\theta_1|\mathbf{y}_1,\mathbf{x}_1)-h(\theta_1|\mathbf{y}_2,\mathbf{x}_2,\mathbf{y}_1,\mathbf{x}_1,\theta_2) \\ &= h(\theta_1|\mathbf{y}_1,\mathbf{x}_1)-h(\theta_1|\mathbf{y}_1,\mathbf{x}_1,\theta_2) \\ &= I(\theta_1;\theta_2|\mathbf{y}_1,\mathbf{x}_1) \\ &= h(\theta_2|\mathbf{y}_1,\mathbf{x}_1)-h(\theta_2|\mathbf{y}_1,\mathbf{x}_1,\theta_1) \\ &= h(\theta_2|e^{j\theta_1}(s_1+z_1),s_1)-h(\theta_2|\theta_1) \\ &\leq \max_{\xi\geq\xi_0} h(\theta_2|e^{j\theta_1}(\xi+z_1))-h(\theta_2|\theta_1) \\ &= h(\theta_2|e^{j\theta_1}(\xi_0+z_1))-h(\theta_2|\theta_1). \end{aligned} \tag{3.33}$$

As claimed, the RHS of (3.33) can be made arbitrarily close to zero by choosing ρ_0 in (3.22) sufficiently large. □

3.1.2 Simulation results

In this section, we numerically evaluate the upper bound $U(\rho)$ in (3.4) and the asymptotic capacity expression (3.16) – the $o(1)$ term is neglected – for a single-antenna Wiener phase-noise channel with standard deviation of the phase-noise increment equal to 6° (Figure 3.1.2) and 20° (Figure 3.1.2). In the figures, we also show the capacity

$$C_{\text{awgn}}(\rho) \triangleq \log(1+\rho) \tag{3.34}$$

of an AWGN channel with SNR equal to ρ, which is a tight upper bound on $C(\rho)$ at low SNR. We also display

$$\tilde{U}(\rho) \triangleq \min\bigl\{C_{\text{awgn}}(\rho), U(\rho)\bigr\}. \tag{3.35}$$

The information rates achievable using QAM constellations of different cardinality, which are lower bounds on $C(\rho)$, are also depicted. We evaluate these rates using the algorithm for the computation of the information rates for finite-state channels described in Chapter 2. Specifically, we use 200 levels for the discretization of the phase-noise process, and a block of 2000 channel uses.

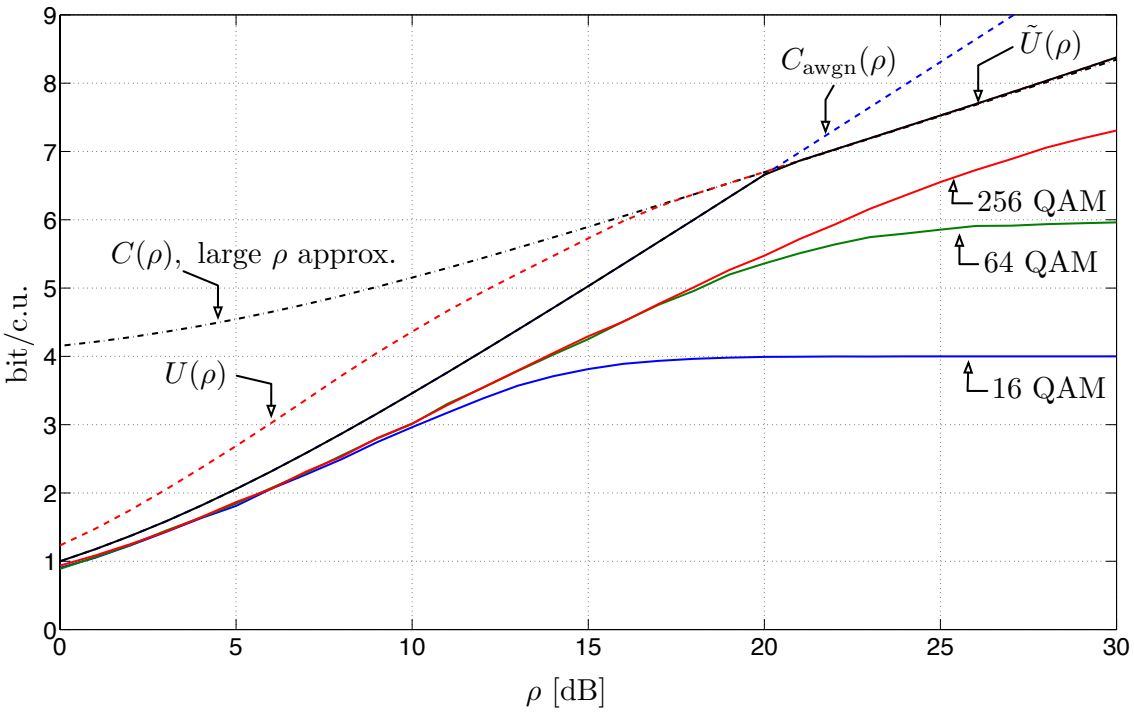

Figure 3.1. The upper bound $U(\rho)$ in (3.4), the asymptotic capacity approximation (3.16), the AWGN capacity (3.34), the tighter upper bound $\tilde{U}(\rho)$ in (3.35), and the rates achievable with 16, 64, and 256 QAM. In the figure, $\sigma_\Delta = 6°$.

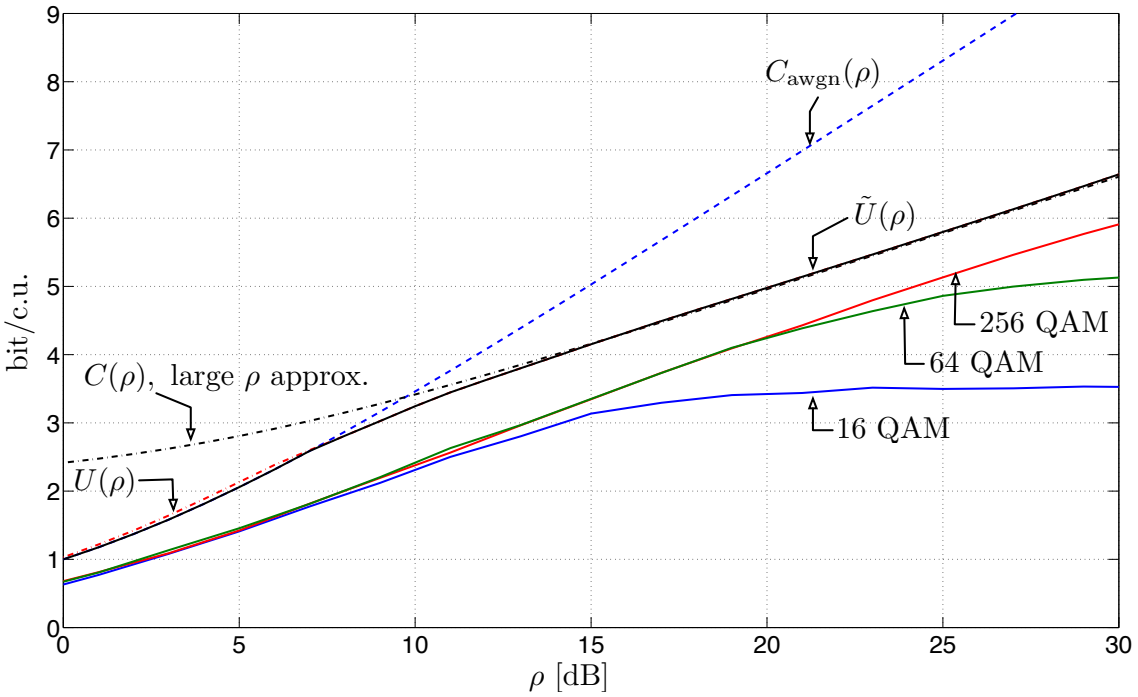

Figure 3.2. The upper bound $U(\rho)$ in (3.4), the asymptotic capacity approximation (3.16), the AWGN capacity (3.34), the tighter upper bound $\tilde{U}(\rho)$ in (3.35), and the rates achievable with 16, 64, and 256 QAM. In the figure, $\sigma_\Delta = 20°$.

In both scenarios C_{awgn} is a tighter upper bound than $U(\rho)$ at low SNR values where the additive noise is the main impairment. In the high-SNR regime, however, $U(\rho)$ is tighter. For medium/high SNR values, the large-SNR capacity approximation (3.16) follows $U(\rho)$ closely. In this regime, QAM constellations incur a penalty of more than $3dB$. The gap to the capacity upper bound might be reduced by replacing QAM with suitably optimized constellations.

3.2 Upper bound with peak energy constraint

We show now a modified version of upper bound presented in Section 3.1 and a closed-form high-SNR expression for the capacity of multiple-antenna systems affected by Wiener phase noise (see Section 1.2.1). Our results are developed again for the scenario where a single oscillator drives all the radio-frequency circuitries at each transceiver (common oscillator setup), the input signal is subject to a peak-power constraint, and the channel matrix is deterministic. This scenario is relevant for LoS multiple-antenna microwave backhaul links with sufficiently small antenna spacing at the transceivers (see Section 1.3).

For the 2×2 multiple-antenna case, for a Wiener phase-noise process with standard deviation equal to 6°, and at the medium/high SNR values at which microwave backhaul links operate, the upper bound reported here exhibits a 3 dB gap from a lower bound obtained using 64-QAM. Furthermore, in this SNR regime the closed-form high-SNR expression is shown to be accurate.

A fundamental way to characterize the impact of phase noise on the throughput of microwave backhaul links is to study their Shannon capacity. Unfortunately, the capacity of the phase-noise channel is not known in closed form even for simple channel models, although capacity bounds and asymptotic results in the limiting regime of high SNR have been reported in the literature.

We present a non-asymptotic capacity upper bound for the case of Wiener phase noise and the practically relevant scenario when the transmit codewords are subject to a peak-power constraint, which is more stringent than the average-power constraint analyzed so far in the phase-noise literature. This upper bound improves on the one recently reported in [2], which was derived under the assumption of codewords subject to an average-power constraint. When particularized to constant-modulus constellations and to single-antenna systems, our bound recovers the upper bound obtained in [1].

We compare our upper bound with lower bounds obtained by evaluating numerically the information rates achievable with QAM constellations. For the case of a Wiener phase-noise process with standard deviation of the phase increments equal to 6°, the gap between our upper bound and the information rates achievable with 64-QAM is about $3dB$ for medium/high SNR values.

We also provide a capacity characterization in the high-SNR regime that is accurate up to a term that vanishes as SNR grow large. This characterization yields a capacity approximation that turns out to be accurate already at moderate SNR values.

We consider the same channel model described in (3.1), i.e. a $M \times M$ MIMO phase-noise channel with memory

$$\mathbf{y}_k = e^{j\theta_k}\mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad k = 1,2,\dots,n. \tag{3.36}$$

Here, $\mathbf{x}_k$ denotes the M-dimensional input vector at discrete time k; $\mathbf{H}$ is the MIMO channel matrix, which we assume deterministic, full-rank, and known to the transmitter and the receiver; $\{\theta_k\}$ is the phase-noise process; and $\{\mathbf{w}_k\}$ is the additive Gaussian noise, which we assume i.i.d. circularly symmetric with zero mean and covariance matrix $\mathbf{I}_M$, i.e., $\mathbf{w}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. We assume again that the phase-noise samples $\{\theta_k\}$ form a Wiener process (see Section 1.2.1).

Peak-amplitude constraint The results proposed in Section 3.1 were derived under the assumption that each transmit codeword $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is subject to the average-power constraint (3.3).

In practice, each codeword entry $\mathbf{x}_k$ must obey a given *peak-power constraint* to avoid distortions due to nonlinearities and saturation effects at the high-power amplifier [27].

To account for this and obtain capacity results that are more relevant in practice, in this paper we substitute (3.3) with the more stringent peak-power constraint

$$\|\mathbf{x}_k\|^2 \leq \rho, \quad k = 1, \ldots, n. \tag{3.37}$$

The peak-power constraint (3.37) has been considered previously in the information-theoretic literature, but not in the contest of phase-noise channels.

Smith [28] proved that the capacity-achieving distribution of an AWGN channel subject to (3.37) is discrete with a finite number of mass point (in contrast, the capacity-achieving distribution under (3.3) [14, Chapter 9] is Gaussian). More recently, Lapidoth [29] characterized the high-SNR capacity of single-antenna stationary fading channels subject to (3.37) in the setting where no *a priori* channel-state information is available at the receiver.

Channel Capacity We are interested in computing the capacity of the MIMO phase-noise channel (3.36), which – under the peak-amplitude constraint (3.37) – is given by

$$C(\rho) = \lim_{n\to\infty} \frac{1}{n} \sup I(\mathbf{y}^n; \mathbf{x}^n) \tag{3.38}$$

where $\mathbf{x}^n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and, similarly, $\mathbf{y}^n = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$. Here, the supremum is over all probability distributions on $\mathbf{x}^n$ that satisfy (3.37) with probability one (w.p.1). In Section 3.2.1, we analyze $C(\rho)$ for the case of $\mathbf{H}$ being unitary. The general full-rank case will be discussed in 3.2.4.

3.2.1 The Unitary Case

Capacity Upper Bound We next present an upper bound on $C(\rho)$ that improves on the one reported in Section 3.1 and [2] for the average-power constrained case.

With some minor adjustments, the bound turns out to be tight in the high-SNR regime.

Before presenting our upper bound, two observations are in order.

i) As $\mathbf{H}$ is known to transmitter and receiver, $C(\rho)$ depends on $\mathbf{H}$ only through its singular values. Since $\mathbf{H}$ is unitary, all singular values are equal to one. Hence, we can (and will) assume without loss of generality that $\mathbf{H} = \mathbf{I}_M$.

ii) In the following proposition, we establish that the capacity-achieving input process $\{\mathbf{x}_k\}$ can be assumed isotropically distributed, a property that will be useful in our analysis.

Proposition 2. *The input process* $\{\mathbf{x}_k\}$ *that achieves the capacity of the channel* (3.36) *when* $\mathbf{H}$ *is unitary can be assumed isotropically distributed. Specifically, if* $\{\mathbf{x}_k\}$ *achieves* $C(\rho)$ *in* (3.38)*, then* $\{\mathbf{U}_k\mathbf{x}_k\}$*, where the matrix-valued random process* $\{\mathbf{U}_k\}$ *is i.i.d. and each* $\mathbf{U}_k$ *is uniformly distributed on the set of* $M \times M$ *unitary matrices, achieves* $C(\rho)$ *as well.*

Proof. The proof, which exploits that $\mathbf{U}_k\mathbf{w}_k \sim \mathbf{w}_k$, follows the same steps as the proof of [25, Proposition 7]. □

Our upper bound on $C(\rho)$ is constructed by extending to the MIMO case the method used in [21] to derive an asymptotic bound on the capacity of stationary single-antenna phase-noise channels. We also use the approach proposed in [23, 24] to make the bound non-asymptotic, and some of the tools developed in [30] to account for the presence of the peak-power constraint (3.37).

For convenience, we introduce the following notation: for every $a > 0$, we let

$$\phi_l(a) \triangleq \underline{/1 + z_l/\sqrt{a}} \tag{3.39}$$

where the random variables $\{z_l\}_{l\in mathbbZ}$ are i.i.d. $\mathcal{N}(0,1)$-distributed and $\underline{/x}$ denotes the phase of the complex number x.

Roughly speaking, $\phi_l(a)$ is the noise level in the estimation of the phase-noise sample θ_l from the channel output $\mathbf{y}_l$ given that the input vector $\mathbf{x}_l$ is known and $\|\mathbf{x}_l\|^2 = a$.

Theorem 3. *The capacity of the channel* (3.36) *under the peak-power constraint* (3.37) *can be upper-bounded as* $C(\rho) \leq U(\rho)$*, where*[1]

$$U(\rho) \triangleq \min_{\alpha>0}\left\{\alpha \log\frac{\rho+M}{\alpha} + d_\alpha + \log(2\pi) + \max_{0\leq\xi\leq\sqrt{\rho}} g_\alpha(\xi,\rho)\right\}. \tag{3.40}$$

Here,

$$\begin{aligned} g_\alpha(\xi,\rho) \triangleq (M-\alpha)\mathbb{E}\left[\log\left(|\xi+z_1|^2 + \sum_{j=2}^{M}|z_j|^2\right)\right] \\ + \alpha\frac{\xi^2+M}{\rho+M} - \sim (|\xi+z_0|^2) \\ - \sim \left(\theta_0 + \phi_0(\xi^2)|\left\{\theta_l+\phi_l(\rho)\right\}_{l=-\infty}^{-1}, |\xi+z_0|\right) \end{aligned} \tag{3.41}$$

where $\{z_l\}$ *are i.i.d.* $\mathcal{N}(0,1)$*-distributed random variables and*

$$d_\alpha \triangleq \log\frac{\Gamma(\alpha)}{\Gamma(M)} - M + 1 \tag{3.42}$$

with $\Gamma(\cdot)$ *standing for the Gamma function.*

[1] *Throughout the work,* log *stands for the natural logarithm.*

Proof. Because of Proposition 1, we can restrict ourselves to isotropically distributed input processes. Specifically, we consider $\{\mathbf{x}_k\}$ of the form $\{\mathbf{x}_k = s_k\mathbf{v}_k\}$, where $s_k = \|\mathbf{x}_k\|$ and $\mathbf{v}_k = \mathbf{x}_k/s_k$, with $\mathbf{v}_k$ uniformly distributed on the unit sphere in $\mathcal{C}^M$ and independent of s_k. We start by using chain rule as follows

$$I(\mathbf{y}^n;\mathbf{x}^n) = \sum_{k=1}^{n} I(\mathbf{y}_k;\mathbf{x}^n|\mathbf{y}^{k-1}). \tag{3.43}$$

By proceeding similarly to [2, Equation (10)], but accounting for the peak-power constraint,[2] we next upper-bound each term on the right-hand side (RHS) of (3.43). We first note that

$$\begin{aligned} I(\mathbf{y}_k;\mathbf{x}^n|\mathbf{y}^{k-1}) &= h(\mathbf{y}_k|\mathbf{y}^{k-1}) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^n) \\ &\overset{(a)}{\leq} h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^n) \\ &\overset{(b)}{=} h(\mathbf{y}_k) - h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k). \end{aligned} \tag{3.44}$$

Here, in (a) we used that conditioning reduces entropy, and in (b) that $\mathbf{y}_k$ and $(\mathbf{x}_{k+1},\ldots,\mathbf{x}_n)$ are conditionally independent given $(\mathbf{y}^{k-1},\mathbf{x}^k)$. We next focus on the conditional differential entropy (the second term) on the RHS of (3.44). Intuitively, the past inputs $\mathbf{x}^{k-1}$ and the past outputs $\mathbf{y}^{k-1}$ can be used to obtain noisy estimates of the past phase-noise samples $\{\theta_l\}_{l=1}^{k-1}$. These estimates help us to guess the value of the current phase-noise sample θ_k. We next use this intuition to obtain a lower bound on $h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k)$, and, hence, an upper bound on $I(\mathbf{y}_k;\mathbf{x}^n|\mathbf{y}^{k-1})$ in (3.43).

For each pair $(\mathbf{y}_l,\mathbf{x}_l)$, $l = 1,\ldots,k-1$, we compute the phase of the projection of $\mathbf{y}_l$ onto $\mathbf{x}_l$. This projection is distributed as $\theta_l + \phi_l(s_l^2)$. Since $\mathbf{y}_k$ and $\mathbf{y}^{k-1}$ are conditionally independent given both $\mathbf{x}^k$ and $\{\theta_l + \phi_l(s_l^2)\}_{l=1}^{k-1}$, we obtain

$$\begin{aligned} h(\mathbf{y}_k|\mathbf{y}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k) &= h\left(\mathbf{y}_k|\left\{\theta_l + \phi_l(s_l^2)\right\}_{l=1}^{k-1},\mathbf{x}^{k-1},\mathbf{x}_k\right) \\ &\geq h\left(\mathbf{y}_k|\left\{\theta_l + \phi_l(\rho)\right\}_{l=1}^{k-1},\mathbf{x}_k\right). \end{aligned} \tag{3.45}$$

In the last step, we used that the best noisy estimate of the past phase-noise samples $\{\theta_l\}_{l=1}^{k-1}$ is achieved by transmitting inputs at peak power, i.e., $s_l^2 = \rho$, $l = 1,\ldots,k-1$.

[2] Recall the the upper bound developed in [2] holds for the average-power constraint case.

Substituting (3.45) into (3.44), we obtain

$$\begin{aligned} I(\mathbf{y}_k; \mathbf{x}^n | \mathbf{y}^{k-1}) &\leq h(\mathbf{y}_k) - h\left(\mathbf{y}_k | \left\{\theta_l + \phi_l(\rho)\right\}_{l=1}^{k-1}, \mathbf{x}_k\right) \\ &= I\left(\mathbf{y}_k; \left\{\theta_l + \phi_l(\rho)\right\}_{l=1}^{k-1}, \mathbf{x}_k\right) \\ &\overset{(a)}{=} I\left(\mathbf{y}_0; \left\{\theta_l + \phi_l(\rho)\right\}_{l=-(k-1)}^{-1}, \mathbf{x}_0\right) \\ &\leq I\left(\mathbf{y}_0; \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1}, \mathbf{x}_0\right) \\ &= I(\mathbf{y}_0; \mathbf{x}_0) + I\left(\mathbf{y}_0; \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} | \mathbf{x}_0\right). \end{aligned} \tag{3.46}$$

Here, (a) follows because $\{\theta_k\}$ is a stationary process. Substituting (3.46) into (3.7) and then (3.43) into (3.38), we obtain

$$C(\rho) \leq \sup\left\{I(\mathbf{y}_0; \mathbf{x}_0) + I\left(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} | \mathbf{x}_0\right)\right\}. \tag{3.47}$$

The supremum in (3.47) is over all probability distributions on $\mathbf{x}_0 = s_0\mathbf{v}_0$ such that s_0 and $\mathbf{v}_0$ are independent, $\mathbf{v}_0$ is uniformly distributed on the unit sphere in $\mathcal{C}^M$, and $s_0^2 \leq \rho$ w.p.1.

We next upper-bound the first term on the RHS of (3.47), which corresponds to the mutual information of a memoryless phase-noise channel with uniform phase noise using a method similar to the one used in [24, 23].

Specifically, we use the duality approach [26, Theorem 5.1] and choose an output probability distribution for which $\mathbf{y}_0$ is isotropically distributed and $r = \|\mathbf{y}_0\|^2$ follows a Gamma distribution with parameters α to be optimized later and $\beta \triangleq (\rho + M)/\alpha$. To summarize the probability density function (pdf) of r is given by

$$q_r(r) = \frac{r^{\alpha-1}e^{-r/\beta}}{\beta^\alpha \Gamma(\alpha)}. \tag{3.48}$$

This output distribution is optimal at high SNR (i.e., it achieves capacity up to a term that vanishes as SNR grows large) for the average-power constraint case [2]. However, it is not optimal for the peak-power constraint case, as we shall discuss in the Appendix A. Nevertheless, it leads to a bound that is accurate for medium SNR values (see 3.2.5).

Using (3.48), we upper-bound $I(\mathbf{x}_0; \mathbf{y}_0)$ as follows (see [2]):

$$\begin{aligned} I(\mathbf{y}_0; \mathbf{x}_0) \leq\ & \alpha \log \frac{\rho + M}{\alpha} + d_\alpha \\ & + (M - \alpha)\left[\log\left(|s_0 + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\ & - h\left(|s_0 + z_0|^2 | s_0\right) + \alpha \frac{\mathbb{E}\left[s_0^2\right] + M}{\rho + M}. \end{aligned} \tag{3.49}$$

Here, d_α is the constant defined in (3.42) and $z_0, z_1, \ldots, z_M$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables.

The second term on the RHS of (3.46) can be evaluated as follows:

$$\begin{aligned} I\left(\mathbf{y}_0; \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} | \mathbf{x}_0\right) &\overset{(a)}{=} I\left(e^{j\theta_0} s_0 + z_0; \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} | s_0\right) \\ &\overset{(b)}{=} I\left(e^{j\theta_0}(s_0 + z_0); \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} | s_0\right) \\ &= I\left(|s_0 + z_0|, \theta_0 + \phi_0(s_0^2); \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} | s_0\right) \\ &\overset{(c)}{=} I\left(\theta_0 + \phi_0(s_0^2); \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1} || s_0 + z_0|, s_0\right) \\ &= h\left(\theta_0 + \phi_0(s_0^2) || s_0 + z_0|, s_0\right) \\ &\quad - h\left(\theta_0 + \phi_0(s_0^2) | \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1}, |s_0 + z_0|, s_0\right) \\ &\overset{(d)}{=} \log(2\pi) \\ &\quad - h\left(\theta_0 + \phi_0(s_0^2) | \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1}, |s_0 + z_0|, s_0\right). \end{aligned} \tag{3.50}$$

Here, (a) follows because $\mathbf{v}_0^H \mathbf{y}_0 \sim e^{j\theta_0} s_0 + z_0$ is a sufficient statistics for $\{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}$; (b) follows because z_0 is circularly symmetric; (c) holds because $|s_0 + z_0|$ and $\{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}$ are independent; finally, (d) holds because $\theta_0 \sim \mathcal{U}[0,2\pi]$.

We substitute (3.49) and (3.50) into (3.47), upper-bound the supremum over all probability distributions on s_0 satisfying $s_0 \leq \sqrt{\rho}$ w.p.1 with the supremum over all deterministic $\xi \in [0, \sqrt{\rho}]$, and tighten the resulting bound by minimizing it over the optimization parameter $\alpha > 0$.

This concludes the proof. □

Remarks The coarser upper bound provided in [2, Th. 2] can be obtained from $U(\rho)$ in (3.40) by assuming perfect knowledge of the past phase-noise samples.

This results in the following cruder lower bound on $h(\mathbf{y}_k | \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k)$ (cf.,(3.45))

$$\begin{aligned} h(\mathbf{y}_k | \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k) &\geq h(\mathbf{y}_k | \theta^{k-1}, \mathbf{x}_k) \\ &= h(\mathbf{y}_k | \theta_{k-1}, \mathbf{x}_k). \end{aligned} \tag{3.51}$$

An even coarser bound can be obtained by assuming perfect knowledge of the additive noise $\phi_0(s_0^2)$ affecting the current phase-noise sample (see (3.50)). This results in the following simple capacity upper bound

$$C(\rho) \leq \sup\left\{I(\mathbf{y}_0; \mathbf{x}_0)\right\} + \log(2\pi) - h(\Delta) \tag{3.52}$$

where Δ is distributed as in(1.3).

The inequality (3.52) can be interpreted as follows: the capacity of a Wiener phase-noise channel is upper-bounded by the capacity of a memoryless phase-noise channel with uniform phase noise, plus a correction term that accounts for the memory in the channel and does not depend on the SNR ρ.

If we now specialize (3.52) to single-antenna systems and we add the additional constraint that $|s| = \sqrt{\rho}$ w.p.1 (which holds, for example, if a PSK constellation is used), the first term on the RHS of (3.52) vanishes and we recover the upper bound previously reported in [1, Theroem 2].

The last term in (3.5) can be computed by using a slightly modified version of the algorithm described in [19].

3.2.2 Asymptotic Behavior

In Theorem 4 below, we present an asymptotic characterization of $C(\rho)$ that generalizes to the MIMO case and to the case of peak-power-constrained inputs the asymptotic characterization reported in [21] for the single-antenna case and average-power-constrained inputs.

Theorem 4. *In the high-SNR regime, the capacity of the Wiener phase-noise channel* (**??**) *behaves as*

$$\begin{aligned} C(\rho) = &\left(M - \frac{1}{2}\right)\log\rho - \log\left(M - \frac{1}{2}\right) - \log\Gamma(M) \\ &+ \frac{1}{2}\log\pi - \left(M - \frac{1}{2}\right) - h(\Delta) + o(1) \end{aligned} \tag{3.53}$$

where $o(1)$ indicates a function of ρ that vanishes in the limit $\rho \to \infty$.

Proof. The proof, which is rather technical, is relegated [2]. □

3.2.3 Average power versus peak power

By comparing the asymptotic capacity expansion provided in Theorem 4 with the one reported in [2, Theorem 3] for the case of average-power-constrained input signals, we can assess the throughput loss at high SNR due to the presence of the more stringent peak-power constraint (3.37). Specifically, let $C_{\text{ap}}(\rho)$ denote the capacity of the channel in (3.36) when the input signal is subject to (3.3) instead of (3.37). Furthermore, let $C(\rho)$ as in (3.2). Then

$$\lim_{\rho\to\infty}\{C_{\text{ap}}(\rho) - C(\rho)\} = \log\Gamma\left(M - \frac{1}{2}\right) - \left(M - \frac{3}{2}\right)\log\frac{1}{M - 1/2} + \left(M - \frac{1}{2}\right). \tag{3.54}$$

For the single-antenna case (i.e., $M = 1$) this asymptotic capacity loss is about 1 bit/s/Hz.

3.2.4 The Non-unitary Case

As mentioned in Section 1.3, practical considerations may force the channel matrix $\mathbf{H}$ to be non-unitary. In this section, we derive non-asymptotic upper and lower bounds on $C(\rho)$ for the general case of full rank $\mathbf{H}$, which are function of the capacity for the case of unitary $\mathbf{H}$. This allows us to extend the results reported in 3.2.1 to the non-unitary case.

Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$, respectively. The following theorem gives upper and lower bounds to $C(\rho)$ for arbitrary full-rank matrix.

Theorem 5. *Let $C_{\text{unitary}}(\rho)$ be the capacity of the channel in (3.36) for the case of unitary $\mathbf{H}$. The capacity for the case of an arbitrary full-rank matrix $\mathbf{H}$ with smallest and largest singular values given by $\sqrt{\lambda_{\min}}$ and $\sqrt{\lambda_{\max}}$, respectively, can be bounded as follows:*

$$C_{\text{unitary}}(\lambda_{\min}\rho) \leq C(\rho) \leq C_{\text{unitary}}(\lambda_{\max}\rho). \tag{3.55}$$

Proof. Since $\mathbf{H}$ has full rank and is known at both sides, precoding at the transmitter can be done in order to invert the channel. Precisely, set $\mathbf{x}_k = \mathbf{H}^{-1}\tilde{\mathbf{x}}_k$, so that (3.36) becomes

$$\mathbf{y}_k = e^{j\theta_k}\tilde{\mathbf{x}}_k + \mathbf{w}_k, \quad k = 1,2,\ldots,n. \tag{3.56}$$

The peak-power constraint (3.37) forces $\tilde{\mathbf{x}}_k$ within the hyperellipsoid

$$\tilde{\mathbf{x}}_k^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1} \tilde{\mathbf{x}}_k \leq \rho, \quad k = 1,\ldots,n \tag{3.57}$$

w.p.1. By definition:

$$C_{\text{unitary}}(\lambda_{\max}\rho) = \lim_{n\to\infty} \frac{1}{n} \sup I(\tilde{\mathbf{x}}^n; \mathbf{y}^n) \tag{3.58}$$

where the supremum is over all distributions on $\tilde{\mathbf{x}}^n$ that satisfy

$$\|\tilde{\mathbf{x}}_k\|^2 \leq \lambda_{\max}\rho, \quad k = 1,\ldots,n, \quad w.p.1. \tag{3.59}$$

The peak-power constraint in (3.59) is looser than (3.57). Indeed,

$$\|\tilde{\mathbf{x}}_k\|^2/\lambda_{\max} \leq \tilde{\mathbf{x}^H}_k \left(\mathbf{H}\mathbf{H}^H\right)^{-1} \tilde{\mathbf{x}}_k \leq \|\tilde{\mathbf{x}}_k\|^2/\lambda_{\min}. \tag{3.60}$$

Hence, if (3.57) holds, then (3.59) holds as well. This implies that

$$C(\rho) \leq C_{\text{unitary}}(\lambda_{\max}\rho). \tag{3.61}$$

In the same way, (3.60) and the definition of $C_{\text{unitary}}(\rho)$ allow us to conclude that

$$C(\rho) \geq C_{\text{unitary}}(\lambda_{\min}\rho). \tag{3.62}$$

□

As a consequence of (3.55), bounds for the case of unitary $\mathbf{H}$ can be transformed into bounds for the case of full-rank $\mathbf{H}$ at the cost of a power offset. Using the asymptotic expression for $C_{\text{unitary}}(\rho)$ reported in (3.53), we see that in the high-SNR regime the gap between the upper and lower bounds in (3.55) is equal to

$$C_{\text{unitary}}(\lambda_{\max}\rho) - C_{\text{unitary}}(\lambda_{\min}\rho) = \left(M - \frac{1}{2}\right) \log \frac{\lambda_{\max}}{\lambda_{\min}} + o(1) \tag{3.63}$$

which tends to a constant as SNR increases.

Note that both the upper and the lower bound are obtained by neglecting the actual structure of (3.57). In order to take this structure into account, a non-isotropic distribution of $\tilde{\mathbf{x}}_k$, with power allocated according to a waterfilling strategy, may result in a tighter lower bound.

3.2.5 Simulation Results

In this section, we numerically compute the upper bound in (3.40) and compare it with the asymptotic expression in (3.53), for a standard deviation of the phase-noise increments equal to $\sigma_\Delta = 6°$, in the two cases $M = 1$ (single-antenna system) and $M = 2$. In Figure 3.3, the curves for the single-antenna case are displayed. The bound approaches the asymptotic expression as SNR grows large, although it remains below it for all the SNR values considered. In the figure, we also show the upper bound from [2, Theorem 2] and its asymptotic expansion [2, Equation (17)]. Although these results were derived for an average-power constraint, they serve as upper bounds for the capacity under a peak-power constraint, since the latter is more stringent than the former. Finally, we also plot an upper bound that is obtained from (3.40) by substituting the conditional differential entropy

$$h\left(\theta_0 + \phi_0(\xi^2) | \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1}, |\xi + z_0|\right) \tag{3.64}$$

with

$$h\left(\Delta + \phi_0(\xi^2) || \xi + z_0|\right). \tag{3.65}$$

This bound, which we refer to as $U_s(\rho)$ (where the letter "s" stands for "simplified") is much simpler to evaluate numerically than $U(\rho)$. Furthermore, its computational complexity does not scale with the number of antennas (on the contrary, the computational complexity of $U(\rho)$ increases exponentially with the number of antennas). Unfortunately,$U_s(\rho)$ is less tight than $U(\rho)$ because

$$h\left(\theta_0 + \phi_0(\xi^2) | \left\{\theta_l + \phi_l(\rho)\right\}_{l=-\infty}^{-1}, |\xi + z_0|\right) \geq h\left(\Delta + \phi_0(\xi^2) || \xi + z_0|\right). \tag{3.66}$$

The newly derived bounds improve on the previous ones by 6-7 dB at moderate and high SNR values, in accordance with what reported in Subsection 3.2.3.

Finally, the numerically computed mutual information for the case of 64-QAM is also shown.[3] A gap ranging from 2 dB to 3 dB is observed between (3.40) and the 64-QAM curve, depending on the SNR. From the plot, we see that the asymptotic capacity expression, which, differently from both upper and lower bounds, is trivial to compute, accurately describes the behavior of the capacity for SNR larger than $16dB$.

In Fig. 3.4, the curves for the case $M = 2$ are shown. As in Fig. 3.3, we also depict the upper bound from [2, Theorem 2] together with its asymptotic version [2, Equation (17)], the simplified upper bound $U_s(\rho)$, and the mutual information achieved by 64-QAM. The newly derived bounds improve on the previous ones by about 3 dB in the high-SNR region. For the MIMO case, the gap between (3.40) and the QAM curve is about 3.5 dB in the high-SNR region and larger for smaller SNR values. In this case, the asymptotic capacity expression seems to describe accurately the capacity behavior for SNR values as small as 4 dB.

It is appropriate to point out that there is no guarantee that our upper bound $U(\rho)$ converges to the asymptotic capacity expression (3.16) as SNR grow large. In fact, the output distribution used in the duality step in the two cases is different. Obtaining a tighter non-asymptotic bound based on the output distribution that is optimal asymptotically remains an open problem.

Plots for the case of a non-unitary matrix $\mathbf{H}$ can be obtained directly from Figure 3.4 by shifting the upper bound to the left by $\lambda_{\max}$ (expressed in dB), and shifting the lower bound to the right by $\lambda_{\min}$ (expressed in dB). The gap between the resulting upper and lower bounds increases proportionally to the logarithm of the ratio between $\lambda_{\max}$ and $\lambda_{\min}$ in accordance to (3.63).

3.2.6 Conclusions

We presented an asymptotic (high-SNR) characterization, as well as nonasymptotic bounds, on the capacity of MIMO microwave backhaul links affected by Wiener phase noise. Our results are developed for the case of common oscillator at the transceivers, and under the practically relevant assumption that the transmit signal is subject to a peak-power constraint. By numerical simulations, we showed that our asymptotic capacity expression, which—differently from the capacity upper and lower bounds—is trivial to compute, is accurate at the SNR values typically encountered in microwave backhaul links (15 dB or higher). In the regime where our asymptotic capacity formula is tight, QAM constellations exhibit a gap of about 3

[3]Specifically, we use the algorithm for the computation of the information rates for finite-state channels proposed [**?**]. We choose 200 levels for the discretization of the phase-noise process, and average over a block of 2000 channel uses.

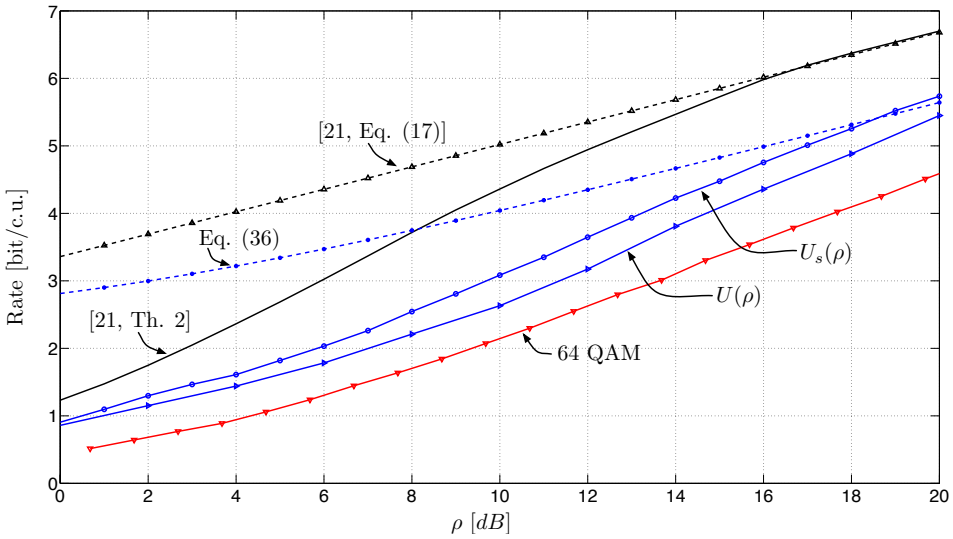

Figure 3.3. The upper bound $U(\rho)$ in (3.40), its simplified version $U_s(\rho)$, the asymptotic capacity approximation (3.53), the upper bound from [2, Theorem 2] and its asymptotic version [2, Equation (17)], and the rates achievable with 64-QAM. In the figure, $\sigma_\Delta = 6°$.

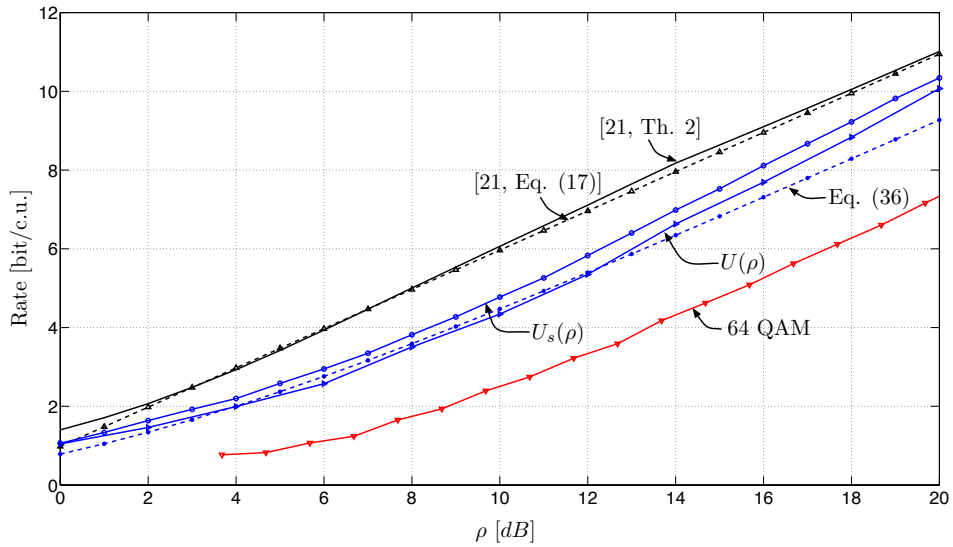

Figure 3.4. The upper bound $U(\rho)$ in (3.40), its simplified version $U_s(\rho)$, the asymptotic capacity approximation (3.53), the upper bound from [2, Theorem 2] and its asymptotic version [2, Equation (17)], and the rates achievable with 64-QAM. In the figure, $\sigma_\Delta = 6°$.

dB. The gap to the capacity upper bound may be reduced by replacing QAM with suitably optimized constellations. Furthermore, the upper bound could be further tightened by substituting the Gamma distribution (3.48) used in the duality step with the asymptotically optimal output distribution (A.20), and then by optimizing over the parameter ϵ. This, however, may further increase the computational complexity associated to the numerical evaluation of the upper bound.

Chapter 4

Phase recovery receivers

4.1 PLL-based receiver for SISO channels

For single-antenna systems, there is a vast literature concerning phase-noise detection algorithms. For the classical phase-locked loop (PLL) solution, see [11] and references therein. More modern solutions rely on message-passing algorithms, among which [31], [32] and [33] are a few examples. Regarding phase noise, the main problem lies in a suitable parameterization of phase-noise messages, a problem which is solved by resorting to Tikhonov distributions in [31] and to particle filtering discretization in [32]. In [33], instead, the Expectation-Maximization algorithm, represented as an algorithm of message-passing, is proposed in order to perform joint phase recovery and channel decoding.

In this Section we present a PLL-based joint detection-decoding scheme, designed to exploit the decision made by an iterative decoder that are fed back to the carrier synchronizer for a SISO channel impaired by a correlated phase noise and in Section 4.2 it will be extended to a LoS MIMO channel. We first present, in Section 4.1.1, a simplified Feed-Forward version of the algorithm without any feedback (we call it Open-Loop scheme) and finally, in Section 4.1.2, we show the advantages to exploit the decision made by the iterative decoder to improve the performance of the detector and the overall system.

4.1.1 Open Loop Code-Aided algorithm

The system model of the SISO channel is

$$y_k = x_k e^{j\theta_k} + w_k, \quad k = 1,2,\ldots,n \tag{4.1}$$

where x_k and y_k are the transmitted and received symbols, respectively, at time instant k, w_k is white Gaussian noise and θ_k is the phase noise sample at instant k,

defined as the sum of phase noise at transmitter and at receiver. Phase noise has been generated using an infinite impulse response (IIR)-filtered Gaussian process, which has been obtained by matching the phase noise mask in Figure 1.1. The transfer function of the filter in Z-transform domain, is given by

$$H(z) = 10^{-3} \frac{0.2932 - 1.137z^{-1} + 1.654z^{-2} + 1.068z^{-3} + 0.2589z^{-4}}{1 - 4.922z^{-1} + 9.693z^{-2} - 9.542z^{-3} + 4.696z^{-4} - 9.245z^{-5}} \tag{4.2}$$

We assume that the transmitted stream is divided into slots of N_d modulated symbols and the beginning of each slot is composed by N_p pilots, while the remaining $N_d - N_p$ symbols are left for the data. The overhead is then obtained as $\eta = N_p/N_d$.

The overhead is not considered in the evaluation of the E_b/N_0 in following results, in order to see only the performance degradation due to the carrier recovery algorithm. The actual information energy to noise ratio $E_{b,inf}/N_0$ is related to the coded energy-to-noise ratio by the following expression

$$\frac{E_{b,code}}{N_0} = R_c(1-\eta)\frac{E_{b,inf}}{N_0} = \frac{1}{m}\frac{E_s}{N_0} \tag{4.3}$$

where R_c is the Forward Error Correction (FEC) code rate, m is the number of bits per modulated symbol, E_s is the energy per modulation symbols, and $E_{b,code}$ is the energy associated to a coded bit.

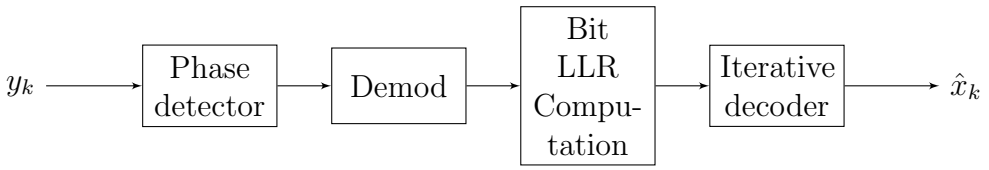

Figure 4.1. Open Loop receiver block diagram

In Figure 4.1 the block scheme of the receiving chain for the Open Loop scenario is presented. The received symbols from the channel are fed into the proposed PLL-based synchronization scheme (it will be discussed later in detail) returning the phase compensated symbols ready to be processed by the soft modulator. A LLR computation block transforms the LLR on symbols in LLR on coded bits that are sent to the iterative decoder (for simulation results we take into account a LDPC and SCCC decoder). Finally decoded bits are provided.

In the following we concentrate on the phase detector, whose structure in the Open-Loop structure is shown in Figure 4.2. The received samples are first split into pilots and data symbols by a switch. The pilot samples enter a Maximum-Likelihood instantaneous phase estimator, followed by a smoothing filter. Such phase estimations based on pilot symbols are linearly interpolated to obtain a rough

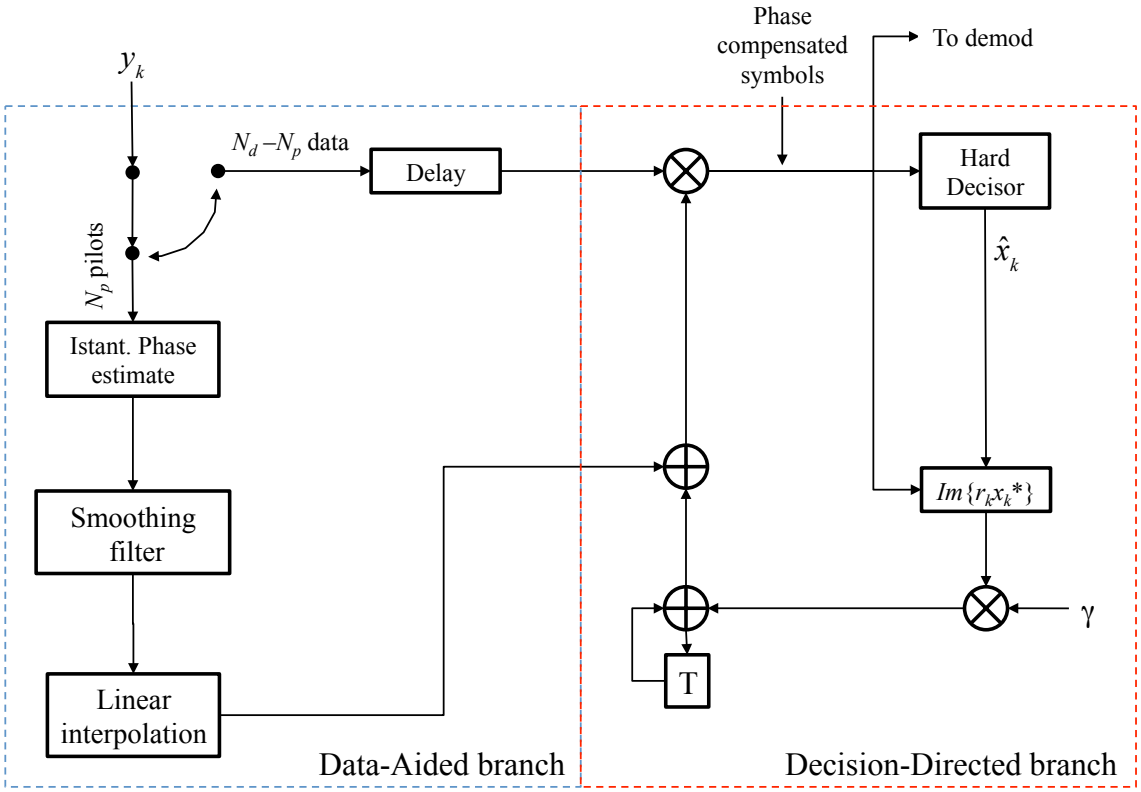

Figure 4.2. Block diagram of the Open-Loop version of PLL-based phase detector

estimation at data rate and added to the one tracked by the Costas Loop of PLL (red branch in Figure 4.2), characterized by the time constant γ [11] to compensate the phase of the received data samples.

Smoothing of phase pilot estimates In order to reduce the burst length without increasing the pilot overhead, and thus reducing the effect of error propagation and cycle slips, we have considered a phase estimation based on distributed individual pilot symbols using an optimal smoothing filter to improve the quality of the estimates. The coefficients of the optimal Minimum Mean Square Error (MMSE) smoothing filter that process the instantaneous phase estimate can be obtained by the Wiener-Hopf equation

$$\mathbf{h} = \left(\mathbf{R}_{\hat{\theta}}^{-1}\mathbf{r}_{\theta\hat{\theta}}\right)^* \tag{4.4}$$

where $\mathbf{R}_{\hat{\theta}}$ is the autocorrelation matrix of the sequence of estimate, and $\mathbf{r}_{\theta\hat{\theta}}$ is the correlation of the estimated phase sample $\hat{\theta}$ at pilot time instant k with the true phase θ at time $k-q$ (q is the delay introduced by the filter, in terms of number of symbols). The optimal smoothing filter is a function of the level of noise and

its length must be such as to include all significant coefficients. We have designed a 7-tap filter, optimized for the phase noise masks in Figure 1.1 and Figure 1.2, a coded signal-to-noise ratio of 13 dB, and a delay of three periods. It is symmetric with the tap coefficients

$$\mathbf{h} = \left[1.09e^{-2}, 6.99e^{-2}, 1.92e^{-1}, 4.543e^{-1}, 1.92e^{-1}, 6.99e^{-2}, 1.09e^{-2}\right].$$

After the instantaneous phase estimation then we use a linear interpolation between successive phase estimates according to

$$\hat{\theta}'_{kN_d+i} = \hat{\theta}_k + \frac{i}{N_d}\left(\hat{\theta}_{k+1} - \hat{\theta}_k\right) \qquad i = 0, \ldots, N_d - 1 \tag{4.5}$$

The interpolation requires to wait for the next pilot, and thus introduces a delay equal to the pilot insertion period.

Costas Loop In Figure 4.2 the red branch of the proposes phase detector shows a Costas Loop [11] running to track the residual phase between the interpolated estimates. Increasing the bandwidth of the loop increases its tracking capabilities but also increases the probability of having cycle slips and error propagation. To find the best trade off between these two effects, we have optimized the normalized Costas Loop bandwidth obtaining a value of 0.0175, corresponding to the value of $\gamma = 0.07$. The residual phase noise at the output of the phase recovery circuit is a highly correlated process owing to the presence of cycle slips and error propagation between successive pilot estimates. The adoption of a memoryless Feed-Forward data aided estimation limits the correlation to the spacing between pilots (N_d). Due to the low signal-to-noise ratio at which the phase synchronizer has to work, the symbol error probability at the output of the hard detector of the Costas Loop is rather high. This makes the loop highly unstable and subject to error propagation and cycle slips for large loop bandwidths. In order to reduce this effect the phase estimation must rely heavily on the pilot symbols. To get a good trade off between the overhead introduced by pilots insertion and the pilot based phase tracking, we keep fixed the overhead introduced by the pilots addition to 5% (i.e. 1 pilot symbol every 20 information symbols).

4.1.2 Iterative Code-Aided algorithm

In this Section we describe an improvement of the Open-Loop algorithm described in Section 4.1.1, which results in a modified version of the algorithm, called Code-Aided recovery algorithm. The motivation behind the design of this scheme is that a more effective phase compensation algorithm may reduce the constraints on the phase mask, reducing the cost of components.

Code-Aided algorithm exploits the information provided by the the soft decision made by an iterative decoder to perform the phase compensation (instead of the embedded hard decsions by Costas Loop in PLL branch of detector) several times, with improved a-priori knowledge of the transmitted symbols. In the following paragraphs we explain the behavior of the algorithm underlining the improvement, in terms of BER, by mean of simulation results.

Code-Aided detection is indicated for high order modulation schemes, e.g. 64-QAM or 256-QAM, more sensitive to phase noise, for which the previous Open-Loop structure may be not sufficient. The feedback between the detector and the decoder not only improves performance of phase synchronization but also the decoder ones exploiting the correlation between bits belonging to the same modulation symbol. The resulting receiver may outperform the Open-Loop structure, even in the absence of phase noise.

In Figure 4.3 the block diagram of the Code-Aided receiver algorithm is depicted. It is similar to the Open-Loop structure with the addition of an outer feedback loops from the decoder to the phase detector. The detector takes the samples from the channel and delivers phase compensated symbols to the demodulator that computes symbols LLRs. The bit LLR Computation block transforms symbols LLRs in bit LLRs that the decoder updates during each iteration. The updated a-priori probabilities are then sent back to the detector, which exploits them to improve the error phase compensation. At the last iteration, the decoder finally returns the decoded data. As you can see in Figure 4.4, the Data-Aided branch (blue part) is the same as the Open-Loop scenario, whereas the Decision-Directed (red part) recovery scheme exploits soft decisions provided by the decoder. The new algorithm requires the optimization of the scheduling between the number of outer (from decoder to detector) and inner (inside the decoder) iterations in order to maximize the performance, keeping constant the total number of decoder iterations (i.e. the product of outer iterations and inner iterations).

The drawback of this solution is the additional memory and complexity required, as the detector and decoder algorithms need to run several times.

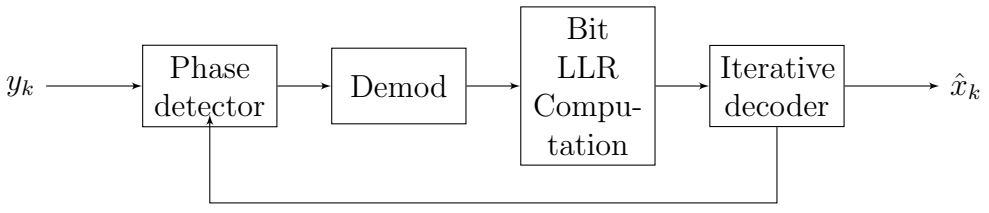

Figure 4.3. Code-Aided receiver block diagram

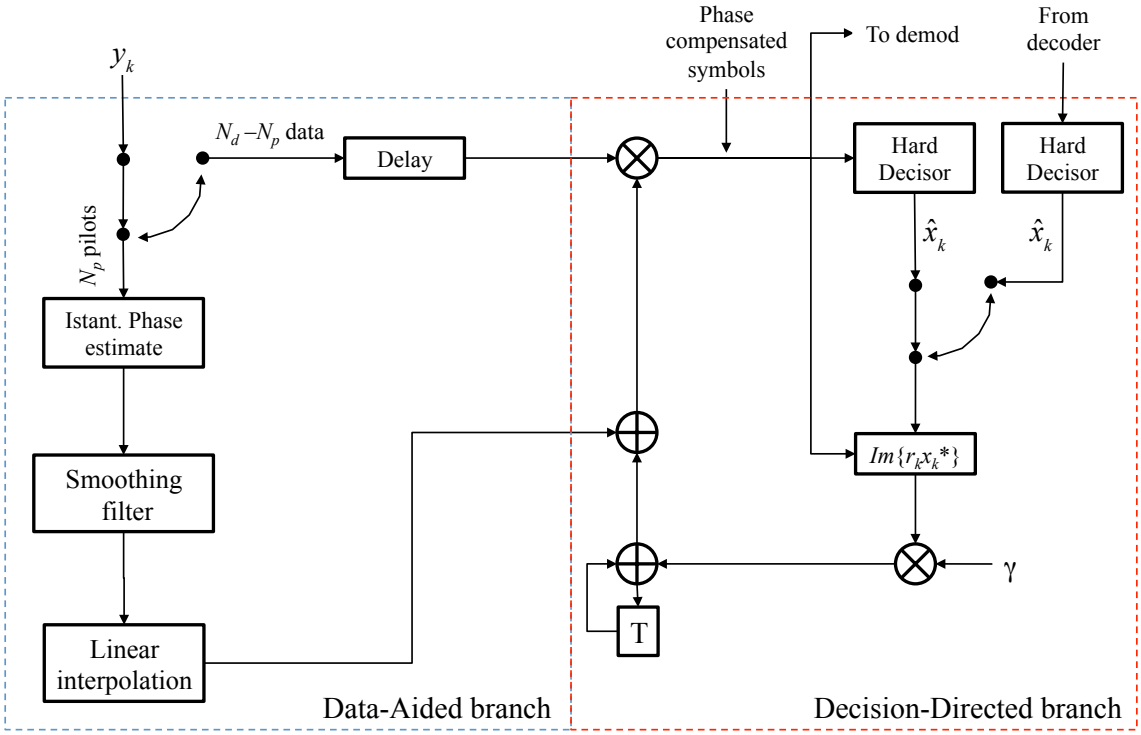

Figure 4.4. Block diagram scheme of proposed Code-Aided PLL-based phase detector for SISO systems

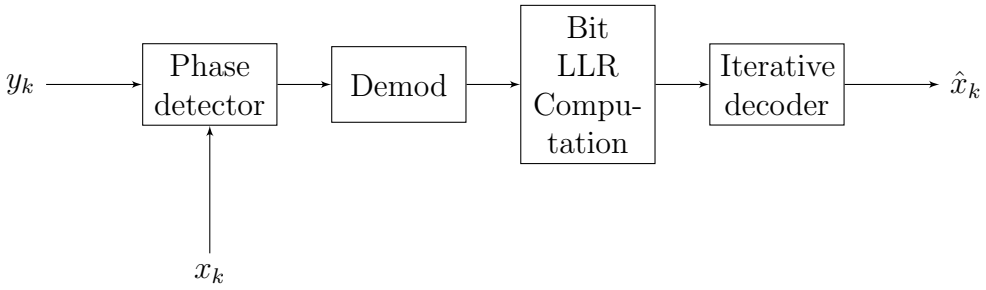

Figure 4.5. Genie-Aided receiver block diagram

Genie-Aided algorithm In Figure 4.5 a so called Genie-Aided version of the Code-Aided algorithm is shown. To evaluate the maximum achievable performance of Code-Aided we simulated a Genie-Aided receiver, for which data are provided with no errors to drive the loop. In other words we assume all symbols to be pilots.

4.1.3 Simulation results

We present in this Section some simulation results for the proposed joint decoder-detector scheme, for 64-QAM and 256-QAM modulations, with the following general setting: we set the symbol rate equal to 6.2 MHz, highly used in back-haul links, and a pilot overhead equal to 1 pilot every 20 symbols (a pilot overhead equal to 5%). For 64-QAM we used a 2 codewords channel interleaver and for 256-QAM modulation we adopted a 4 codewords channel interleaver (the length of interleavers has been optimized by simulation).

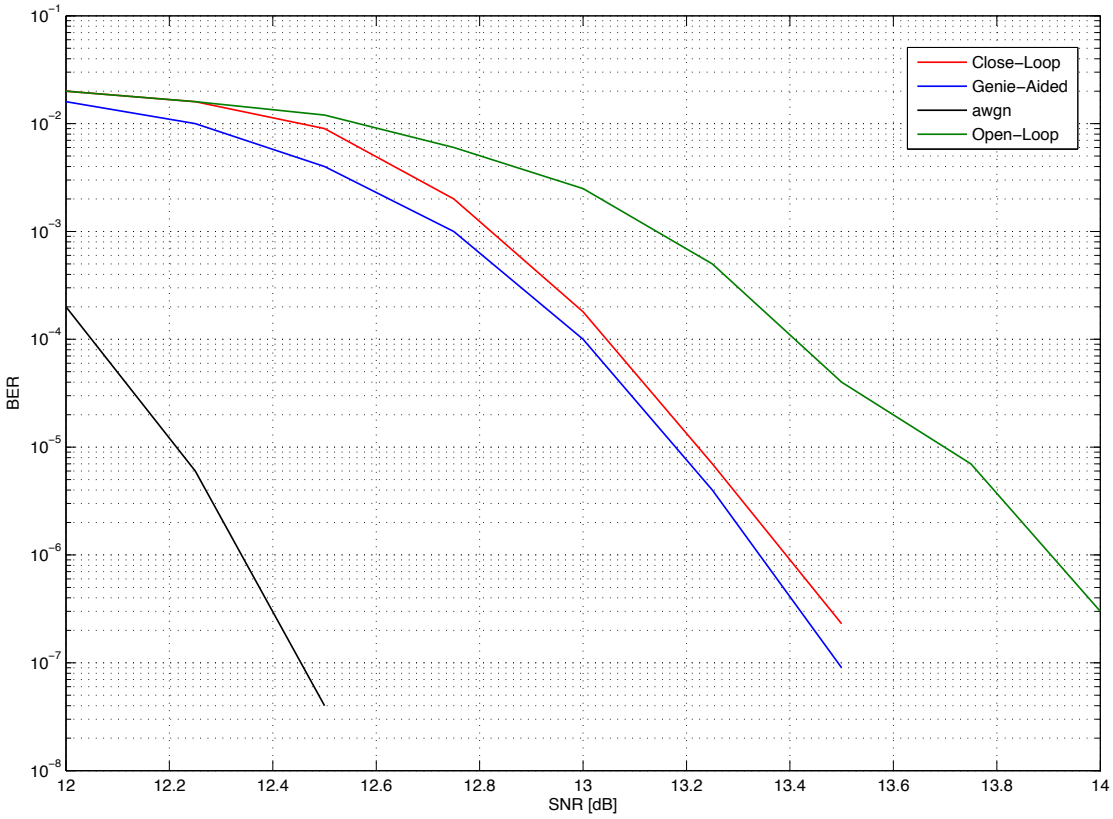

Figure 4.6. Performance, in terms of BER, of Code-Aided, Genie-Aided and Open-Loop algorithms for 64-QAM modulation and LDPC decoder, as well as AWGN reference

In Figure 4.6 and Figure 4.7 we show the simulation results for 64-QAM, with a LDPC and SCCC decoder respectively. In Figure 4.8 the results for LDPC and SCCC are showed together for comparison. Notice that we used 20 iterations inside the LDPC decoder and only 10 iterations for the SCCC decoder because of the more decoding capability of serially concatenated iterative decoders. The feedback iteration settings, i.e. the number of decoder iterations and the number of iterations

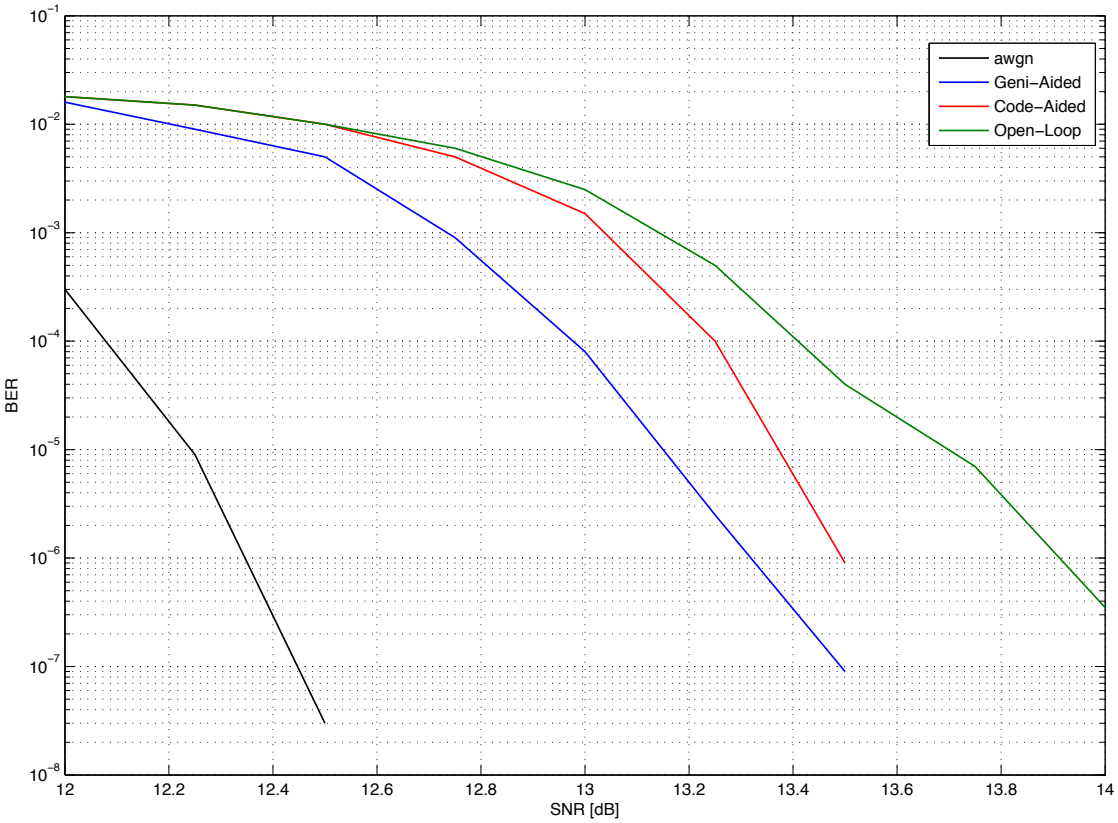

Figure 4.7. Performance, in terms of BER, of Code-Aided, Genie-Aided and Open-Loop algorithms for 64-QAM modulation and SCCC decoder, as well as AWGN reference

between decoder and synchronizer, as well as the PLL time constant γ have so been optimized under those constraints. For LDPC we have $\gamma = 0.1$ with 2 outer iterations (from decoder back to detector) and 10 inner decoder iteration per each; for SCCC we get $\gamma = 0.2$ and again only 2 outer iterations (and consequently 5 decoder iterations) are necessary to get the maximum improvement of performance w.r.t. the Open-Loop solution. In the charts, for the sake of comparison, we have also reported the results obtained using the Open-Loop algorithm under the same PLL settings.

In Figure 4.6 we can see that, for 64-QAM modulation, the activation of the Closed-Loop algorithm permits to reduce the loss w.r.t. Genie-Aided performance from 0.6 dB, related to the Open-Loop scheme, to less than 0.05 dB using a LDPC decoder. In Figure 4.7 the gap is reduced from 0.5 dB for the Open-Loop algorithm to 0.15 dB with the adoption of the Closed-Loop scheme for the SCCC decoder. The BER level for performance analysis is 10^{-6}. In both cases we notice a steeper slope.

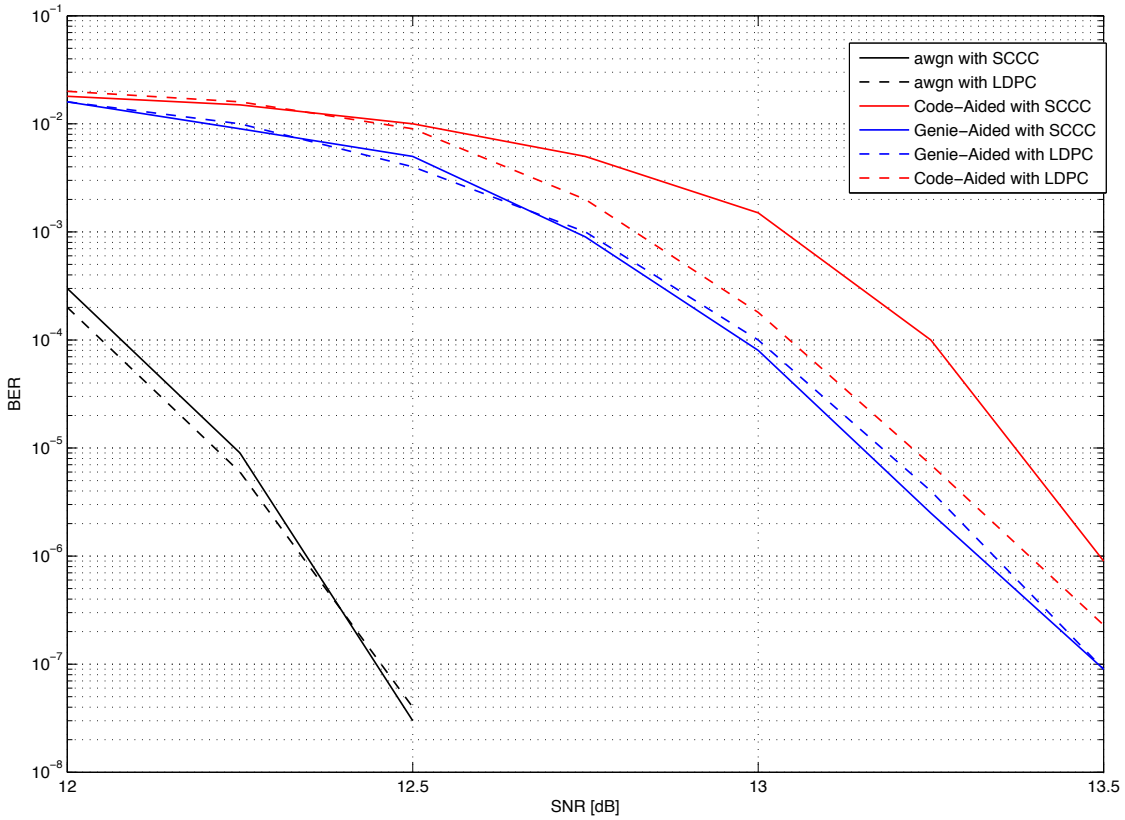

Figure 4.8. Performance, in terms of BER, of Code-Aided, Genie-Aided and Open-Loop algorithms for 64-QAM modulation and both LDPC and SCCC decoders, as well as AWGN reference

The differences between LDPC and SCCC are non-existent for the Genie-Aided as you can clearly see in Figure 4.8.

In Figure 4.9 we show simulation results for 256-QAM, with 4 codewords interleavers. Only the case of SCCC decoder has been considered in this case. The curves show that the gap w.r.t. the Genie-Aided reference has been reduced from 0.6 dB for Open-Loop algorithm to 0.2 dB thank to the adoption of Closed-Loop scheme, with a much steaper slope.

Phase recovery results compared with capacity As we said in Section 1.2.3 the phase noise process can be split into two components, one at low frequency and one at high frequency. The slowly varying phase component is usually well recovered by phase synchronization systems (included the detector presented in this Chapter), so we are interested in comparing the performance of the proposed solution with the theoretical limit of a channel impaired by the residual fast phase noise. In

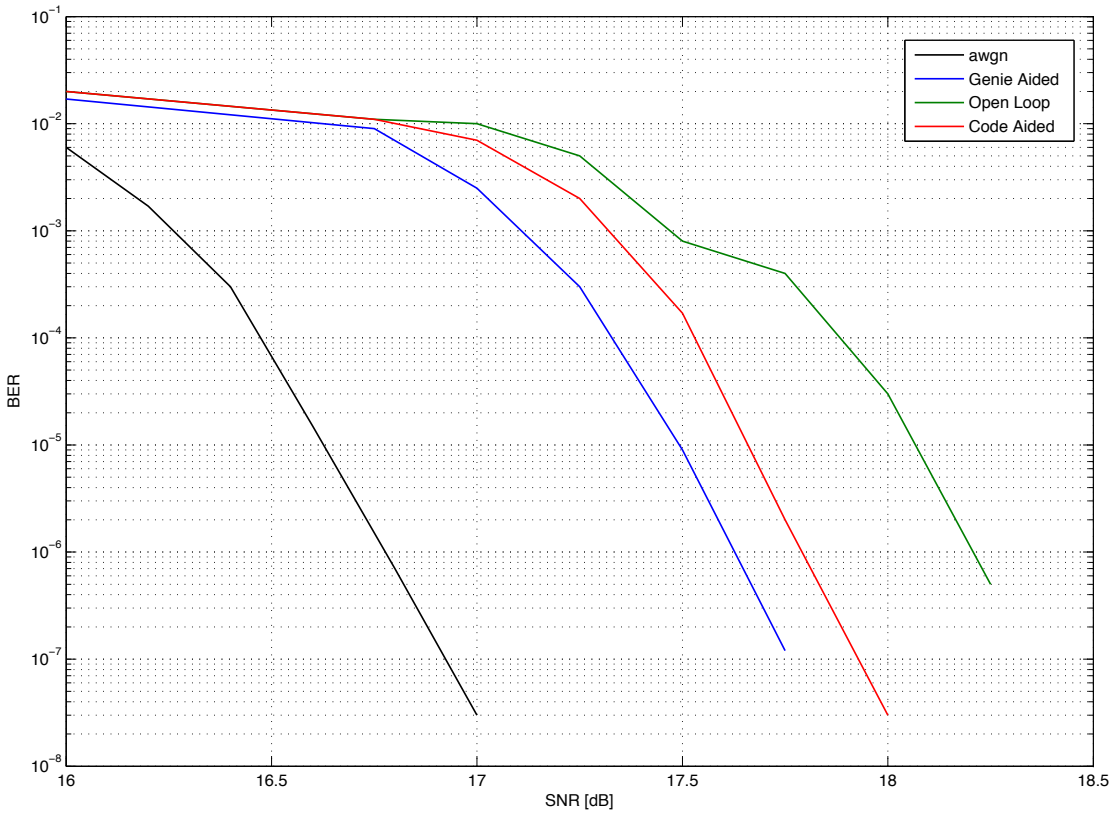

Figure 4.9. Performance, in terms of BER, of Genie-Aided, Code-Aided and Open Loop algorithms, for 256-QAM, as well as AWGN reference

Figure 4.10 and 4.11 blue curve represents the PSD measured by simulation of high frequency phase noise process component, and the red curve is the actual phase noise mask (see Figure 1.1 and Figure 1.2), respectively for 64-QAM and 256-QAM modulations. To let the PSD correctly fits the phase noise mask, we chose (recalling the definition of the AR1 process in (1.5)) $\alpha = 0.998$ and $\sigma_\Delta = 1.64\,\text{deg}$ for 64-QAM and $\alpha = 0.998$ $\sigma_\Delta = 0.4\,\text{deg}$ for 256-QAM modulation. We used those AR1 processes to model the phase noise for the computation of the constraint capacity with the tool described in Chapter 2.

In the following we want to compare the performance of the proposed Closed-Loop phase detector to the theoretical limit expressed by the AWGN capacity and the lower bound to capacity derived in Chapter 2. In Figure 4.12 you can see a chart related to the 64-QAM modulation where the green curve represents the AWGN capacity and the red curve is the capacity for a system impaired by phase noise modeled as we described before; the gap between them, due to the presence of phase noise, is 0.6 dB in correspondence to a target spectral efficiency $\eta = 5.63$

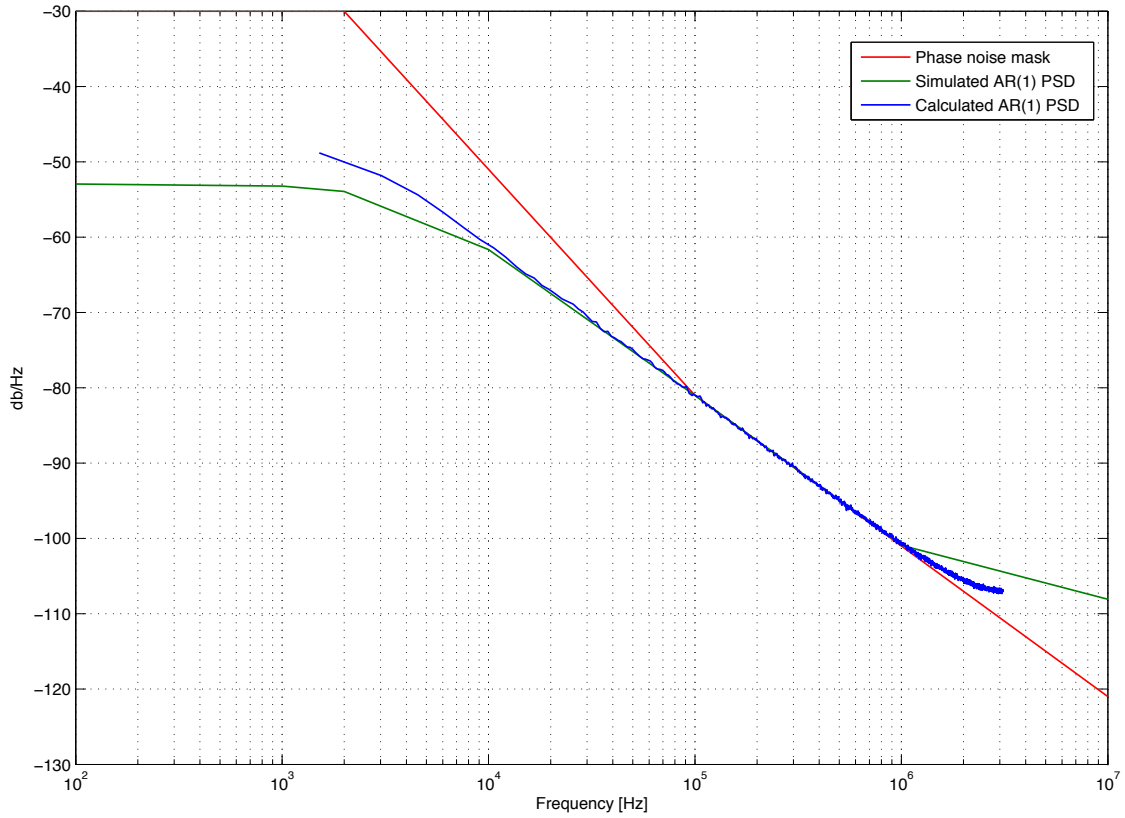

Figure 4.10. Phase noise mask provided by Ericsson, PSD measured and simulated for only high frequency phase noise component for 64-QAM

(dotted black line in the chart), due to the presence of a decoder with a code rate of 15/16. The green mark represents the coded Eb/N0 related to a BER=10^{-6} for the AWGN channel; the loss w.r.t. the AWGN capacity is about 1.1 dB, depending only on thermal noise. The loss of our proposed Closed-Loop scheme w.r.t. the constrained capacity (blue curve) is 1.8 dB (upside down triangle red mark), while the minimum gap we could expect if we used the Genie-Aided algorithm is about 1.6 dB. All those values are taken at a reference BER level of 10^{-6}. In Figure 4.13 similar simulation results for 256-QAM are depicted; also in this case the BER level is 10^{-6} and the loss between the capacity curves is about 0.7 dB, while the performance for the AWGN channel is 1.4 dB far from the AWGN capacity, a loss greater then the previous case. The best performance achieving Genie-Aided algorithm shows a loss of 1.5 dB w.r.t. the constrained capacity, while the Closed-Loop scheme shows a gap w.r.t. capacity of 1.7 dB.

We can conclude that the performance of proposed Closed-Loop scheme are quite close to the channel capacity and that the optimization of feedback settings permits

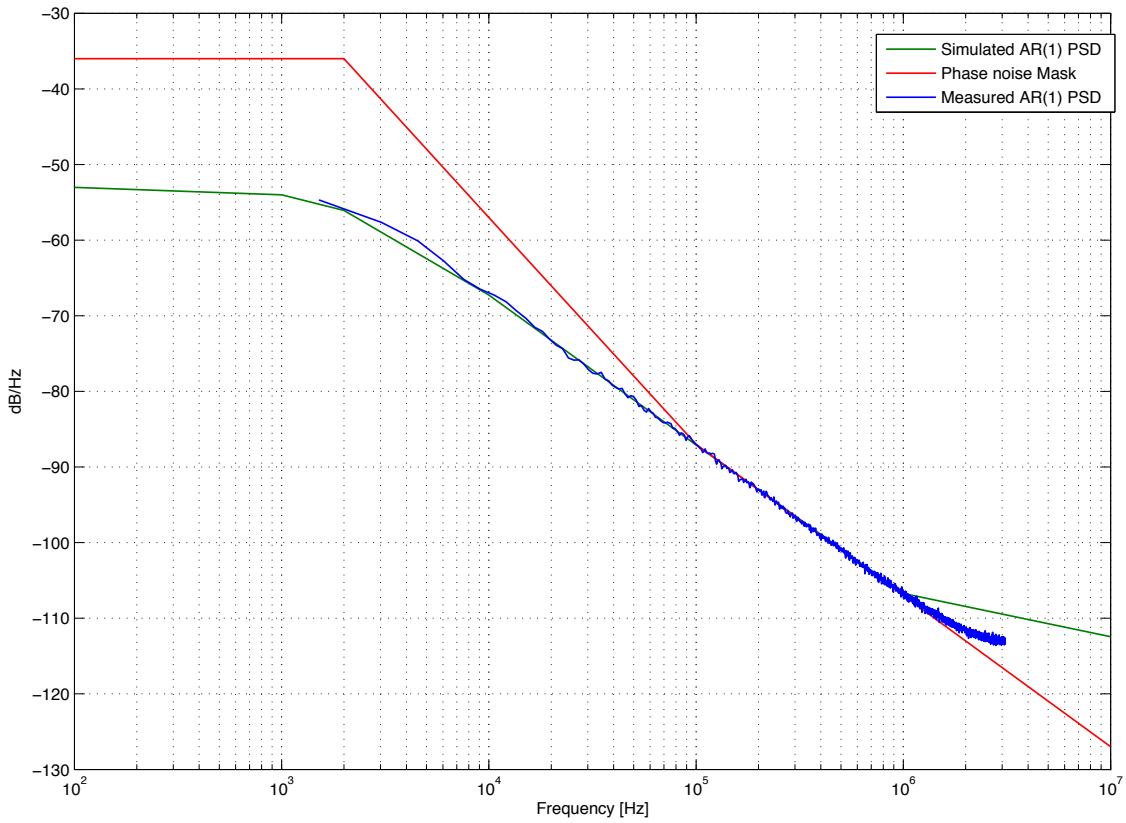

Figure 4.11. Phase noise mask provided by Ericsson, PSD measured and simulated for only high frequency phase noise component for 256-QAM

to get just a loss of about 0.3 dB from the performance of a system considering all symbols to be pilots.

4.2 PLL-based receiver for LoS MIMO channels

In this Section we present a synchronization algorithm for a 2×2 MIMO LoS system, that is an extension of what we described in Section 4.1.2. We assume that the channel matrix is unitary, a condition that can be imposed on the system by suitably arranging the antennas (see Section 1.3). The algorithm is a combination of a Data-Aided (DA) part and a Decision-Directed (DD) part. The former is composed by a pilot-based phase estimator and a filtering stage exploiting the time correlation of phase-noise processes and the space correlation offered by multiple antennas. The latter is based on PLL algorithms to estimate and compensate a set of phase parameters, also based on the feedback provided by a channel decoder.

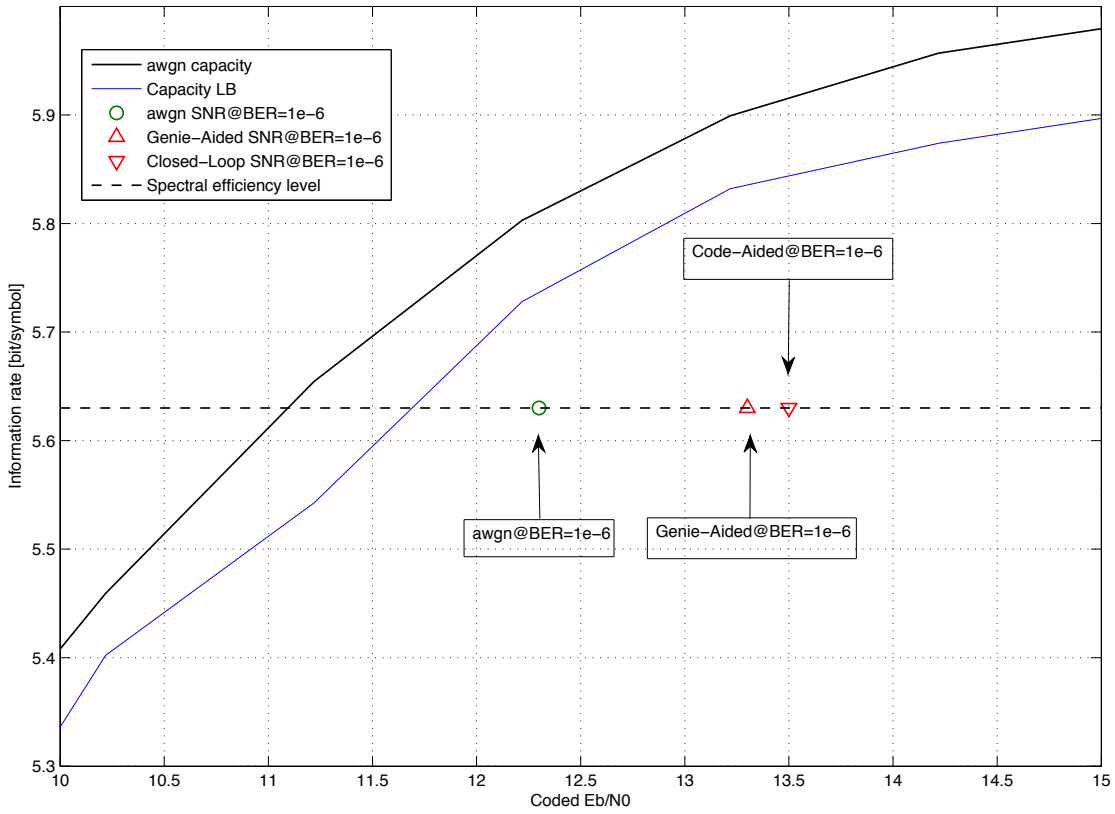

Figure 4.12. AWGN capacity, lower bound to capacity and results achieved by phase detector, compared to AWGN performance, for 64-QAM

While the PLL is a well-known circuit for phase tracking in single antenna systems (see [11, Section 5.3]), we present a non-trivial generalization for MIMO systems that is able to conjugate very good performance with an affordably low complexity.

All the algorithms for phase recovery in SISO systems, when extended in a trivial way to MIMO systems, result in solutions that either perform poorly or have a too high complexity to be practically feasible. Out of the papers that tackle the problem of phase-noise recovery for MIMO systems, in [34] a space-time filter is used to simultaneously exploit the correlation between phase-noise samples in time and between phase processes at different receive antennas. In [35] a phase estimator is presented, which is based on the Least Square and Extended Kalman filter algorithms.

We consider a 2×2 MIMO LoS system impaired by phase noise, whose input-output relationship, at time instant k, is given by

$$\widehat{\mathbf{y}}_k = \widetilde{\mathbf{H}}_k \mathbf{x}_k + \mathbf{n}_k, \tag{4.6}$$

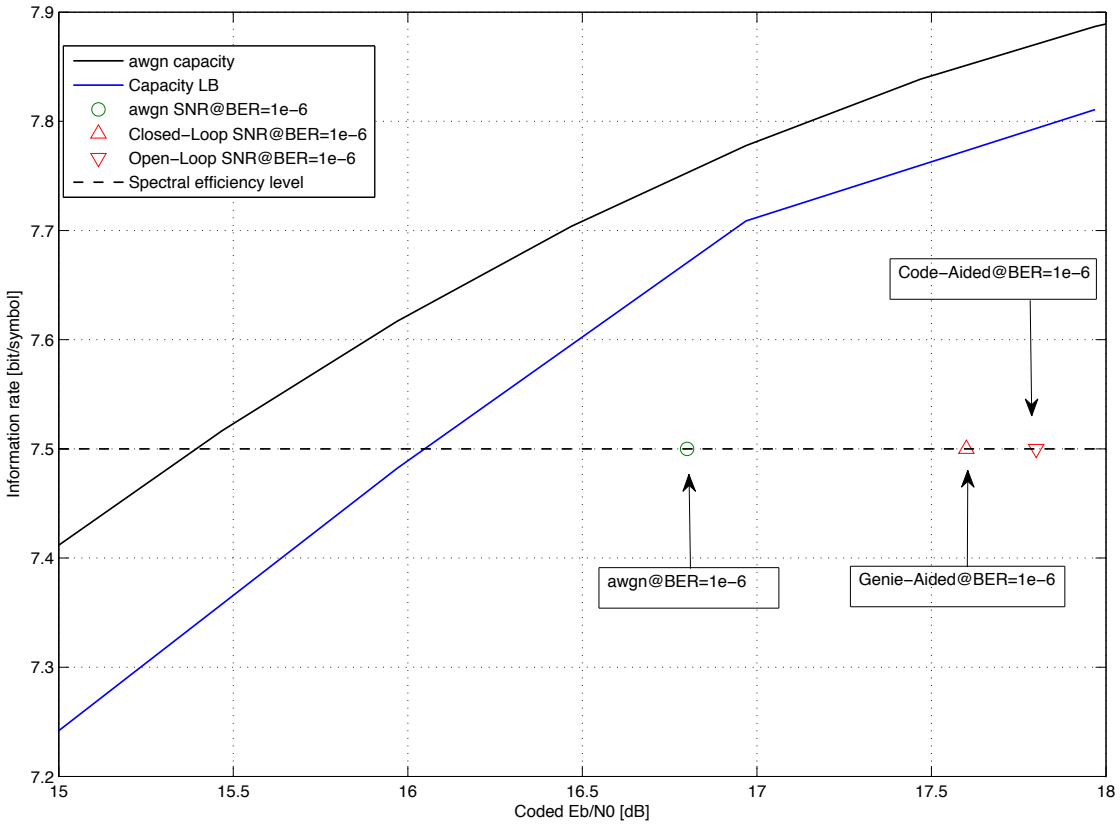

Figure 4.13. AWGN capacity, lower bound to capacity and results achieved by phase detector, compared to AWGN performance, for 256-QAM

where $\mathbf{x}_k$ is the vector of the two transmitted symbols, $\hat{\mathbf{y}}_k$ the vector of the two received samples and $\mathbf{n}_k$ is a vector of uncorrelated, white Gaussian noise samples with zero mean and variance σ^2 per real dimension. The equivalent channel matrix $\widetilde{\mathbf{H}}_k$ is defined as

$$\widetilde{\mathbf{H}}_k = \mathbf{\Theta}_k^r \mathbf{H} \mathbf{\Theta}_k^t. \tag{4.7}$$

where $\mathbf{\Theta}_k^t$ and $\mathbf{\Theta}_k^r$ are the diagonal matrices of phase-noise coefficients, at the transmit and receive side respectively, written as:

$$\mathbf{\Theta}_k^t = \begin{bmatrix} e^{j\theta_{t1}^k} & 0 \\ 0 & e^{j\theta_{t2}^k} \end{bmatrix},$$

$$\mathbf{\Theta}_k^r = \begin{bmatrix} e^{j\theta_{r1}^k} & 0 \\ 0 & e^{j\theta_{r2}^k} \end{bmatrix},$$

where θ_{ti}^k and θ_{rj}^k are the realizations of phase-noise processes at the i-th transmit antenna and j-th receive antenna, respectively, assumed to be unknown at both

sides. Such *atomic* phase-noise processes are assumed in this paper to be independent, identically distributed stationary processes with arbitrary power spectral density $S_\theta(f)$ depicted in Figure 1.1 and generated using the infinite impulse response (IIR)-filtered Gaussian process in (4.2).

The 2×2 channel matrix $\mathbf{H}$ will be considered fixed and known at the receiver. Moreover, throughout the paper, we will also suppose that $\mathbf{H}$ is (proportional to) a unitary matrix, i.e., $\mathbf{H}^H\mathbf{H} = \mathbf{I}$. Such assumption is not unrealistic in the LoS scenario, where, by proper positioning of the transmit and receive antennas, the link gains are controllable by design. In this case, a unitary matrix allows for the maximum channel capacity among all matrices with a given Frobenius norm. A possible parameterization of the 2×2 unitary matrices, which will be useful in the following, reads as:

$$\mathbf{H} = \begin{bmatrix} \cos\alpha & e^{j\psi}\sin\alpha \\ -e^{-j\psi}\sin\alpha & \cos\alpha \end{bmatrix}. \tag{4.8}$$

where $\alpha \in [0, \pi)$ and $\psi \in [0{,}2\pi)$.

Before describing the proposed solution, let us notice that the entries $\widetilde{h}_{mn}$ of the equivalent channel matrix defined in (4.7) are related to the entries h_{mn} of $\mathbf{H}$ as follows (omitting the time index k):

$$\widetilde{h}_{mn} = h_{mn}e^{j\phi_{mn}}, \;\; m, n = 1{,}2. \tag{4.9}$$

where we have defined the sum phase process:

$$\phi_{mn} = \theta_{rm} + \theta_{tn}. \tag{4.10}$$

Notice that, out of the four sum phase processes, one is a linear combination of the other three, namely:

$$\phi_{22} = \phi_{12} + \phi_{21} - \phi_{11}. \tag{4.11}$$

(In the general $N_t \times N_r$ case, out of N_tN_r sum phase processes, only $N_t + N_r - 1$ are linearly independent.) Thus, the receiver has to estimate only three sum phase processes, which are however correlated with each other.

4.2.1 Data-aided detector

The data-aided scheme is represented in Fig. 4.14 by a cascade of functional block, described in details in the following paragraphs. Instantaneous phase estimatation block is followed by an MMSE filter, exploiting spatial correlation of MIMO system, and smoothing filters, exploiting temporal correlation of phase noise and returning the estimates of atomic phase noise processes. These are finally interpolated and recombined to obtain three sum phase-noise parameters to be delivered to the decision-directed part, described in 4.2.2.

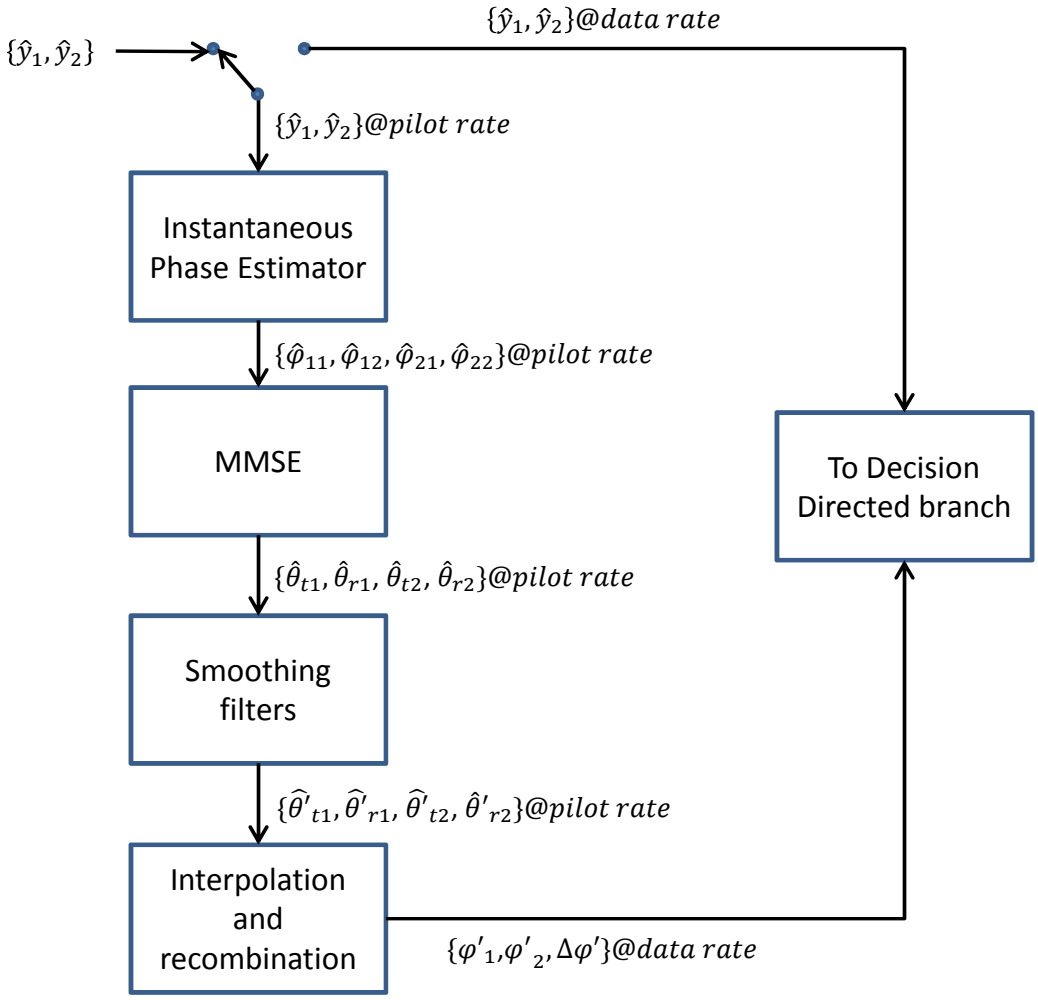

Figure 4.14. Block diagram of Data-Aided branch of proposed solution

Istantaneous phase estimation We suppose that pilots are regularly inserted in the transmission frame, at a rate of one transmitted pilot vector every N_d data transmissions. To simplify the instantaneous phase estimation, an orthogonal (single-antenna) pilot matrix can be used, namely

$$\mathbf{p} = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}. \tag{4.12}$$

This pilot structure allows for trivially obtaining rough estimates $\hat{\phi}_{mn}$ for the four sum phase processes. Notice that such rough estimates are linearly related to the

atomic phase-noise samples:

$$\begin{bmatrix} \hat{\phi}_{11} \\ \hat{\phi}_{12} \\ \hat{\phi}_{21} \\ \hat{\phi}_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_{t1} \\ \theta_{t2} \\ \theta_{r1} \\ \theta_{r2} \end{bmatrix} + \mathbf{n}. \tag{4.13}$$

where $\mathbf{n}$, the estimation noise, can be assumed to be Gaussian with zero mean and variance equal to the channel termal noise variance σ^2. In writing (4.13), we have assumed that phase-noise processes can be considered as constant during pilot transmission time, which is true in most realistic cases.

Minimum Mean Square Error filtering The correlation of the sum phase processes due to MIMO signalling can be exploited to improve the quality of the data-aided estimation through a linear MMSE filter, in order to counteract the effect of noise. The MMSE filter is based on the relationship in (4.13). Assuming that the atomic phase processes are independent and uniformly distributed over [0,2π), a straightforward computation holds the following expression:

$$\begin{bmatrix} \widehat{\theta}_{t1} \\ \widehat{\theta}_{t2} \\ \widehat{\theta}_{r1} \\ \widehat{\theta}_{r2} \end{bmatrix} = C \begin{bmatrix} \epsilon+3 & -1 & \epsilon+3 & -1 \\ -1 & \epsilon+3 & -1 & \epsilon+3 \\ \epsilon+3 & \epsilon+3 & -1 & -1 \\ -1 & -1 & \epsilon+3 & \epsilon+3 \end{bmatrix} \begin{bmatrix} \widehat{\phi}_{11} \\ \widehat{\phi}_{12} \\ \widehat{\phi}_{21} \\ \widehat{\phi}_{22} \end{bmatrix} \tag{4.14}$$

where $C = \frac{1}{(\epsilon+4)(\epsilon+2)}$ and ϵ is the ration between noise variance and phase noise variance:

$$\epsilon = \frac{3\sigma^2}{\pi^2}. \tag{4.15}$$

Smoothing filter and interpolation While the estimated phase-noise samples at the output of the MMSE filter in (4.14) are independent, further enhancement against noise can be obtained by exploiting time correlation in the phase-noise samples. In order to do that, the proposed solution is to smooth the estimated atomic phase-noise processes by a bank of four linear MMSE filters (see [14]), matched to the actual time correlation of such processes, which depends on the power spectral density $S_\theta(f)$. In Figure 4.14, the output (at pilot rate) of the bank of smoothing filters is represented by $\left\{\hat{\theta}'_{t1}, \hat{\theta}'_{r1}, \hat{\theta}'_{t2}, \hat{\theta}'_{r2}\right\}$.

Finally such estimates are interpolated and recombined to obtain a sequence (at data rate) of rough sum-processes phase estimates to be delivered to the decision-directed algorithm. We choose the following variables as the three-element basis for

our solution:

$$\begin{cases} \varphi'_1 = \hat{\theta}_{t1} + \hat{\theta}_{r1} \\ \varphi'_2 = \hat{\theta}_{t2} + \hat{\theta}_{r1} \\ \Delta\varphi' = \hat{\theta}_{r1} - \hat{\theta}_{r1} \end{cases} . \tag{4.16}$$

4.2.2 Decision-directed algorithm

We start by providing some intuition to understand the motivation lying behind this work. Let $\mathbf{y}$ be the vector containing received symbols after to be processed with the Hermitian version of channel matrix

$$\mathbf{y} = \mathbf{H}^{\dagger}\hat{\mathbf{y}} = \mathbf{H}^{\dagger}(\widetilde{\mathbf{H}}\mathbf{x} + \mathbf{n}). \tag{4.17}$$

Expanding the expression of (4.17), we obtain

$$\begin{aligned} y_1 =& x_1 e^{j\varphi_1}(\cos\alpha^2 + \sin\alpha^2 e^{j\Delta\varphi}) + \\ & x_2 e^{j\varphi_2} e^{j\psi} \sin\alpha\cos\alpha(1 - e^{j\Delta\varphi}) + \tilde{n}_1, \\ y_2 =& x_2 e^{j\varphi_2}(\sin\alpha^2 + \cos\alpha^2 e^{j\Delta\varphi}) + \\ & x_1 e^{j\varphi_1} e^{-j\psi} \sin\alpha\cos\alpha(1 - e^{j\Delta\varphi}) + \tilde{n}_2, \end{aligned} \tag{4.18}$$

where $\tilde{n}_1$ and $\tilde{n}_2$ are filtered version of gaussian samples noise. It can be noticed that, if $\Delta\varphi$ was equal to zero, then each received symbols would depend just on the correspondent transmitted symbol. This means that, if we are able to well compensate the differential phase error process at the receiver, then we can estimate the other processes independently.

An other way to understand the idea this work is based on is considering the correlation coefficient c between y_1 and y_2. From (4.17) and assuming that $\Delta\varphi$ is equal to zero, we can write

$$c_k = \mathbb{E}[y_1 y_2^*] \propto \mathbb{E}\left[x_1 x_2^*\right] = 0 \tag{4.19}$$

This result shows again that, if we are able to compensate $\Delta\varphi$, then channels of MIMO system will get orthogonal and could be treated in independent way.

The PLL-based algorithm Considering our model, we can see that there are four atomic phase noise processes, but one of them is a deterministic function of the others. The idea is that we can estimate just three sum-processes to perform a correct phase compensation using one of the atomic parameters as reference. In this way we are able to low down the complexity of overall system and to simplify the treatment. A possible choice of phase-noise parameters array is the following

$$\mathbf{\Phi} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \Delta\varphi \end{bmatrix} = \begin{bmatrix} \theta_{t1} + \theta_{r1} \\ \theta_{t2} + \theta_{r1} \\ \theta_{r2} - \theta_{r1} \end{bmatrix} . \tag{4.20}$$

In this case we have chosen θ_{r1} as reference to reduce the overall complexity of the synchronization stage, obtaining the following equivalent channel model

$$\widetilde{\mathbf{H}} = \begin{bmatrix} 1 & 0 \\ 0 & e^{j\Delta\varphi} \end{bmatrix} \begin{bmatrix} \cos\alpha & e^{j\psi}\sin\alpha \\ -e^{-j\psi}\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} e^{j\varphi_1} & 0 \\ 0 & e^{j\varphi_2} \end{bmatrix}. \tag{4.21}$$

Let's consider now the log-likelihood function Λ of received samples y_1 and y_2

$$\begin{aligned} \Lambda(\mathbf{y}) &= \log\{P(y_1, y_2|\varphi_1, \varphi_2, \Delta\varphi, x_1, x_2)\} \\ &\propto -\frac{1}{2\sigma^2}[|y_1 - \tilde{y}_1|^2 + |y_2 - \tilde{y}_2|^2] \end{aligned} \tag{4.22}$$

In (4.22) y_1 and y_2 are received symbols and $\tilde{y}_1$ and $\tilde{y}_2$ are the hypothesis about them, dependent on phase parameters we want to estimate, i.e. versions of (4.18) without termal noise. To derive the optimal phase paramter estimates we take the maximum of (4.22) w.r.t. the phase parameters

$$(\varphi_1, \varphi_2, \Delta\varphi)_{opt} = \underset{(\varphi_1,\varphi_2,\Delta\varphi)}{\arg\max}\, \Lambda(\mathbf{y}). \tag{4.23}$$

To solve the maximization problem we have to solve the following non-linear system of equations obtained setting to zero the partial derivatives of $\Lambda(\mathbf{y})$ w.r.t each phase parameter

$$\begin{cases} \frac{\partial\lambda(\mathbf{y})}{\partial\varphi_1} = 0 \\ \frac{\partial\lambda(\mathbf{y})}{\partial\varphi_2} = 0 \\ \frac{\partial\lambda(\mathbf{y})}{\partial\Delta\varphi} = 0 \end{cases}. \tag{4.24}$$

Lets consider the partial derivative w.r.t φ_1. We have

$$\begin{aligned} \frac{\partial\lambda(\mathbf{y})}{\partial\varphi_1} &\propto \frac{\partial}{\partial\varphi_1}[-|y_1 - \tilde{y}_1|^2 - |y_2 - \tilde{y}_2|^2] \\ &\overset{(a)}{=} \frac{\partial}{\partial\varphi_1}[-\Delta y_1 \Delta y_1^* - \Delta y_2 \Delta y_2^*] \\ &\overset{(b)}{=} 2\Re\left\{\frac{\partial\tilde{y}_1}{\partial\varphi_1}\Delta y_1^* + \frac{\partial\tilde{y}_2}{\partial\varphi_1}\Delta y_2^*\right\} = 0 \end{aligned} \tag{4.25}$$

where in (a) we define Δy_i as $(y_i - \tilde{y}_i)$, for $i = 1{,}2$ and in (b) we exploit the linearity of derivative operator. The expression in (4.25) takes two solutions (a minimum and a maximum) due to the 2π-periodicity of phase parameters. From (4.25) implies the following expression

$$\arg\left\{-j\frac{\partial\tilde{y}_1}{\partial\varphi_1}\Delta y_1^* - j\frac{\partial\tilde{y}_2}{\partial\varphi_1}\Delta y_2^*\right\} = 0. \tag{4.26}$$

The solution of (4.26) is the following

$$\begin{aligned}
&\arg\left\{-j\frac{\partial\tilde{y}_1}{\partial\varphi_1}\Delta y_1^* - j\frac{\partial\tilde{y}_2}{\partial\varphi_1}\Delta y_2^*\right\} = \\
&= \arg\left\{e^{j\varphi_1}x_1y_1^*(\cos^2\alpha + \sin^2\alpha e^{j\Delta\varphi})\right. \\
&\left.+e^{j\varphi_1}x_1y_2^*e^{-j\psi}\sin\alpha\cos\alpha(1-e^{j\Delta\varphi})\right\} \\
&= \varphi_1 + \arg\left\{x_1y_1^*(\cos^2\alpha + \sin^2\alpha e^{j\Delta\varphi})\right. \\
&\left.+x_1y_2^*e^{-j\psi}\sin\alpha\cos\alpha(1-e^{j\Delta\varphi})\right\} \\
&= \varphi_1 - \arg\left\{x_1^*y_1(\cos^2\alpha + \sin^2\alpha e^{-j\Delta\varphi})\right. \\
&\left.+x_1^*y_2e^{j\psi}\sin\alpha\cos\alpha(1-e^{-j\Delta\varphi})\right\} = 0.
\end{aligned} \tag{4.27}$$

From (4.27) we get the optimal estimate of φ_1

$$\begin{aligned}
\varphi_{1,opt} = \arg\big\{&x_1^*(y_1\cos^2\alpha + y_2e^{j\psi}\sin\alpha\cos\alpha) \\
&+x_1^*e^{-j\Delta\varphi}(y_1\sin^2\alpha - y_2e^{j\psi}\sin\alpha\cos\alpha)\big\}.
\end{aligned} \tag{4.28}$$

The other two solutions of (4.24) are computed in a similar way taking the derivatives of likelihood function w.r.t φ_2 and $\Delta\varphi$. Finally we have

$$\begin{aligned}
\varphi_{1,opt} &= \arg\big\{x_1^*(y_1\cos^2\alpha + y_2e^{j\psi}\sin\alpha\cos\alpha) \\
&\quad +x_1^*e^{-j\Delta\varphi}(y_1\sin^2\alpha - y_2e^{j\psi}\sin\alpha\cos\alpha)\big\} \\
\varphi_{2,opt} &= \arg\big\{x_2^*(y_1e^{-j\psi}\sin\alpha\cos\alpha + y_2\sin^2\alpha) \\
&\quad +x_2^*e^{-j\Delta\varphi}(y_2\cos^2\alpha - y_1e^{-j\psi}\sin\alpha\cos\alpha)\big\} \\
\Delta\varphi_{opt} &= \arg\big\{x_1^*e^{-j\varphi_1}(y_1\sin^2\alpha - y_2e^{j\psi}\sin\alpha\cos\alpha) \\
&\quad +x_2^*e^{-j\varphi_2}(y_2\cos^2\alpha - y_1e^{-j\psi}\sin\alpha\cos\alpha)\big\}.
\end{aligned} \tag{4.29}$$

The solutions are dependent on each other, but under the reasonable assumption, in feedback error domain, that $(\varphi_1, \varphi_2, \Delta\varphi) \simeq 0$, we can approximate (4.29) with the following system of independent error generating functions

$$\begin{aligned}
\delta\varphi_1|_{(\varphi_2,\Delta\varphi)\simeq 0} &= \arg\{y_1x_1^*\} \\
\delta\varphi_2|_{(\varphi_1,\Delta\varphi)\simeq 0} &= \arg\{y_2x_2^*\} \\
\delta\Delta\varphi|_{(\varphi_1,\varphi_2)\simeq 0} &= \arg\big\{x_1^*(y_1\sin^2\alpha - y_2e^{-i\psi}\sin\alpha\cos\alpha) \\
&\quad +x_2^*(y_2\cos^2\alpha - y_1e^{-i\psi}\sin\alpha\cos\alpha)\big\}
\end{aligned} \tag{4.30}$$

The first two equations are the error functions of two independent SISO channels relating just on the respective received and transmitted symbols. The third one is the error generating function for $\Delta\varphi$, depending on all received symbols and the

structure of the channel. As the estimations of φ_1 and φ_2 in closed-loop branch are independent to each other, we can use uncorrelated Costas loop filters. We have, at time instant k

$$\begin{cases} \phi_1^{(k)} = \phi_1^{(k-1)} + \gamma_1 \delta\varphi_1 \\ \phi_2^{(k)} = \phi_2^{(k-1)} + \gamma_2 \delta\varphi_2 \\ \phi_3^{(k)} = \phi_3^{(k-1)} + \gamma_3 \delta\Delta\varphi \end{cases} . \tag{4.31}$$

where (ϕ_1, ϕ_2, ϕ_3) are phase estimates and $(\gamma_1, \gamma_1, \gamma_1)$ are time constants, representing the step-size characteristic of Costas Loop. Finally estimates provided by open loop (4.16) are added to (4.31).

4.2.3 Simulation results

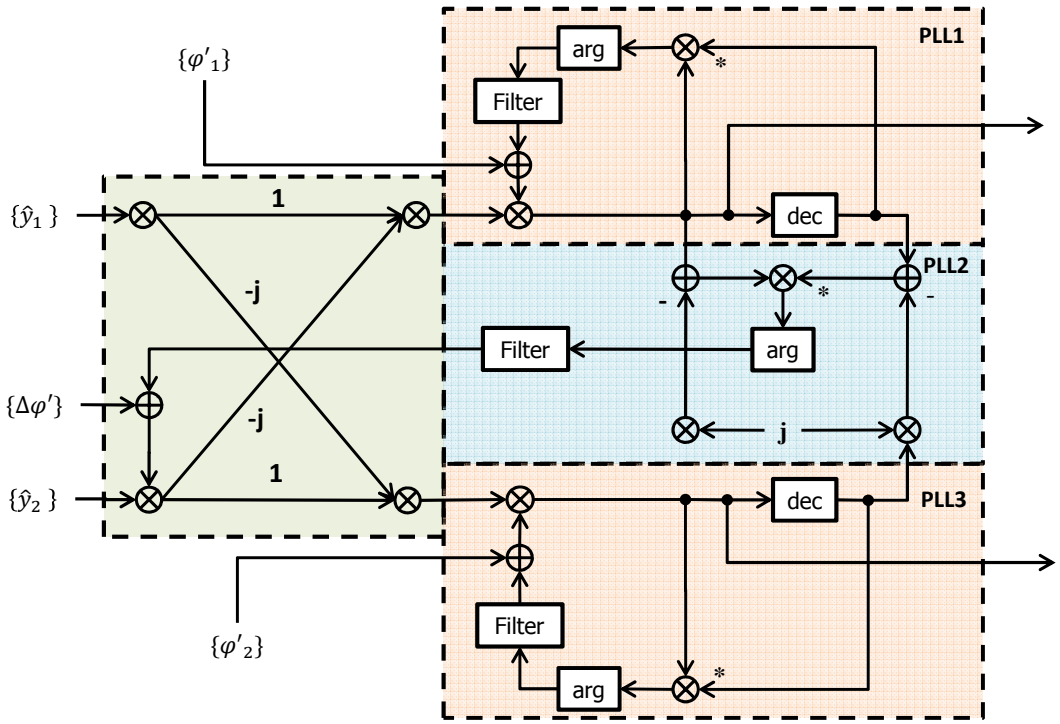

Figure 4.15. Block diagram of Decision-Directed branch of proposed solution

In this Section we present the design of simulations and performances of proposed 2×2 MIMO scheme for 256-QAM modulation, with the following channel model,

obtained setting in (4.21) $\alpha = \frac{\pi}{4}$ and $\psi = \frac{\pi}{2}$

$$\widetilde{\mathbf{H}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & e^{j\Delta\varphi} \end{bmatrix} \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \begin{bmatrix} e^{j\varphi_1} & 0 \\ 0 & e^{j\varphi_2} \end{bmatrix}. \tag{4.32}$$

You can notice that the resulting channel matrix $\mathbf{H}$ is the same presented in (1.12). Regarding the phase noise statistics, phase noise processes at each antenna have been generated from the phase noise mask in Figure 1.2 and generated using the infinite impulse response (IIR)-filtered Gaussian process in (4.2).

In order to have a good trade-off between pilot overhead and synchronization capability, orthogonal pilot symbols are inserted in the transmitted codewords with a rate of $N_p = 1$ pilot every $N = 20$ symbols (i.e. a data frame length of $N_d = 19$), to perform the rough phase estimation in Data-Aided branch. The noise level in the optimum smoothing filter, set by simulation way, is 0.007 with 7 taps.

Transmitted symbols are coded with a Serially Concatenated Convolutional Code (SCCC) with code rate 15/16 and codewords of length 12720 coded bits. The modulation format is 256-QAM and random channel interleaving has been employed. At the receiver, the SCCC decoder performs at most 10 iterations. The PLL-based phase detector has been optimized in sumulating way, obtaining the optimal values for time constants $(\gamma_1, \gamma_2, \gamma_3) = (0.07, 0.08, 0.01)$.

In Figure 4.15 you can see the circuit of described PLLs for the channel model in (4.32). With previous setting, simulation results in Figure 4.16 show the performance, in terms of BER of the proposed algorithm. The black curve refers to the AWGN case, i.e. in absence of phase noise, the blu curve refers to the BER of the algorithm and the red curve refers to a scenario in which we switched off all PLL circuits in Decision-Directed branch. As you can see, the introduction of PLL-based detector achieves a gain of about 1 dB (at BER=10^{-5}) w.r.t the Data-Aided synchronization scheme.

As a remark, we obtained the same results (not reported here) also with the introduction of the optimum filter exploiting temporal and spatial correlation described in [34].

In Figure 4.17 we want to underline the contribution of each functional block in Data-Aided part, showing the Mean Square Error (MSE) of residual phase noise as function of SNR. Green curve depicts the MSE after the istantaneous phase estimator, red curve refers to the MSE resulting from MMSE filter and blu curve represents the behaviour of MSE after smoothing filtering stage. As you can see MSE decreases after each block processing and the more we work in low SNR region, i.e. the more termal noise is high, the more this behavior is evident.

Complexity analysis We now present the computational complexity of proposed solution, analyzed for each block of the algorithm, in terms of number of operations

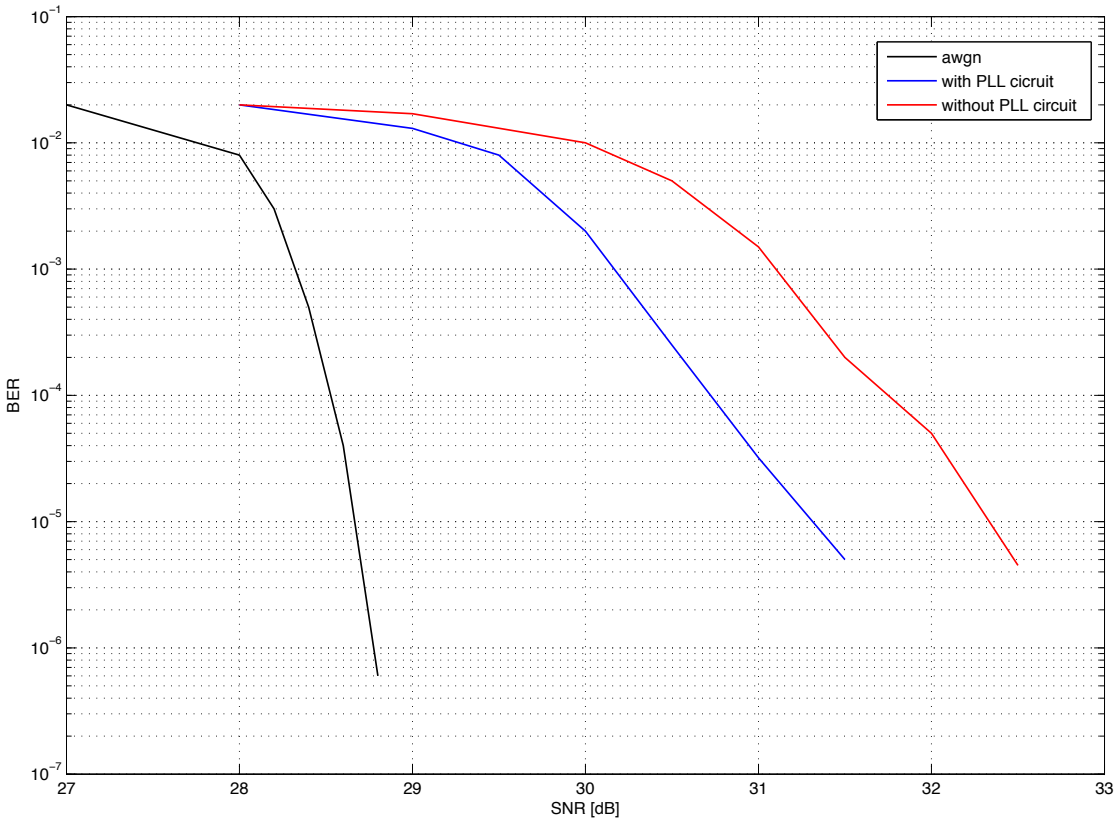

Figure 4.16. Performance, in terms of BER, of proposed phase detector for MIMO channel with and without the PLL circuit and AWGN channel

per information symbol. For each block, we computed the required number of sums, products, divisions and look-up table accesses (LUTs), where the last refers to computations of non-elementary functions like exponentiation, trigonometric functions, etc., that can be realized through the access to a proper look-up table. Regarding sums, products and divisions, they are always considered with respect to real variables (so that one complex sum corresponds to two real sums, etc.). Table 4.1 shows the complexity analysis referring to simulations settings. In view of this table, we can state that the complexity of the proposed solution is very low and feasible for a lot of cases of practical interest, due to the simplicity of PLL filters adopted here.

In this Section, we have introduced a phase-noise estimator for MIMO systems, which is based on PLL algorithm. This work is a generalization of classical approaches for phase detection in SISO systems, but it is far from being straightforward, because multiple antennas introduce overlapping of phase noice samples, not easy to manipulate. Simulation results show that the performance of the proposed

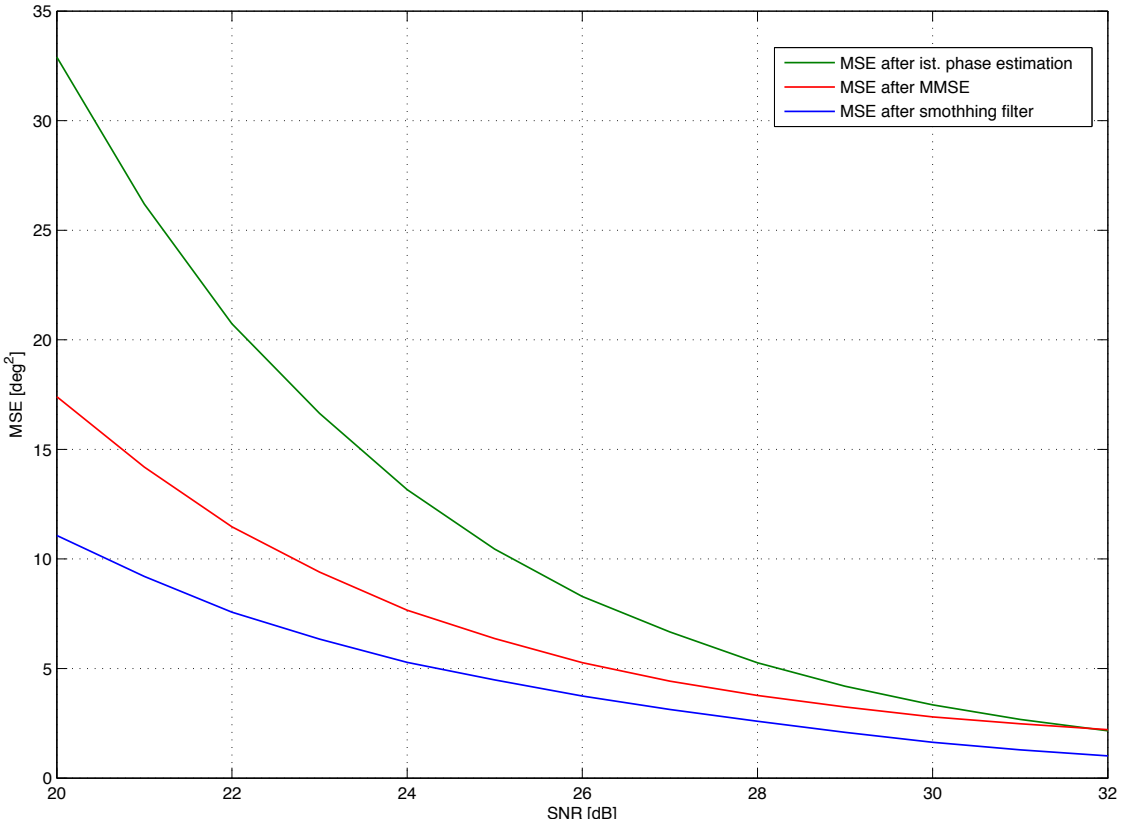

Figure 4.17. MSE of residual phase noise after Istantaneous Phase Estimation block, MMSE filter and Smoothing filter

solution loses about 2.5 dB from the performance of AWGN scenario and its computational complexity is taken very low, due to the simplicity of PLL circuits adopted here.

This solution is so a good trade-off between performance and complecity, making it feasible for real applications.

Variance analysis of phase detection In this Section we present a comparison between the performance of the proposed PLL-based phase detector, in terms of the variance of the estimated phase, and the Cramér-Rao bound (BCRB) derived in [36]. In the following we refer to general settings described in Appendix B where we show the extended computation of the BCRB.

Lets consider a 2×2 MIMO LoS system with channel matrix defined in (1.12) and with independent equal oscillators at each antenna and Wiener phase-noise model. In particular, we consider the case where the blocklength T is large enough and derive analytical expressions for the bound in the middle of the block.

Functional blocks	LUTs	Sums	Products	Divisions
Inst. phase est.	0.21	0.84	1.68	0.21
MMSE filter	0	0.63	0.63	0
Smoothing filter	0	2.21	2.21	0
Interp.	0	4	8	0
Recombination	0	5	0	0
PLL filtering	6	26	31	0
PLL hard decision	4	0	0	0
Total	**10.21**	**38.68**	**43.52**	**0.21**

Table 4.1. Number of operations per transmitted symbol for simulated scenario described in Section 4.2.3

Figure 4.18 shows the BCRB as a function of the time step n for a blocklength $T = 600$. BCRBs for both smoothers (blue curves) and filters (red curves) are shown. Two values of SNR are plotted: SNR = 5 dB (circles) and SNR = 10 dB (crosses), where SNR $= 1/\sigma^2$. Moreover, $\rho = 0.3$ degrees (see Appendix B for further details); although phase noise is not Wiener, the model defined by (B.2) with $\rho = 0.3$ degrees is a good approximation for the actual statistics. As it can be seen, all cases converge to their limit after a transient dependent on initial conditions. The case with low SNR employs more time steps (about 200) to reach its limit value than the case with high SNR, which takes about 100 time steps.

Next, we show the comparison between the BCRBs and the performance of two different phase detectors. The first is the EM-based detector described in [37],[8], whose reference BCRB is the one in (B.14), since it smooths the phase estimates by looking at past and future samples. The second is the PLL-based proposed solution for MIMO channels presented in this Section, and as such must be compared to the BCRB in (B.28).

For the PLL solution, we show the Genie-Aided performance where the data are perfectly known, while for the EM solution, data knowledge is actually obtained by iterating between detector and decoder. As a channel code, we have used a serial concatenation of convolutional codes (SCCC), with rate 15/16 and a codeword length of 12720 coded bits. We employed 4^i-QAM modulation formats, for $i = 1, \ldots, 4$. Pilots are introduced at a rate of 1 pilot every 20 transmitted symbols. Random interleaving has been employed before transmission. All the other parameters are as in Figure 4.18.

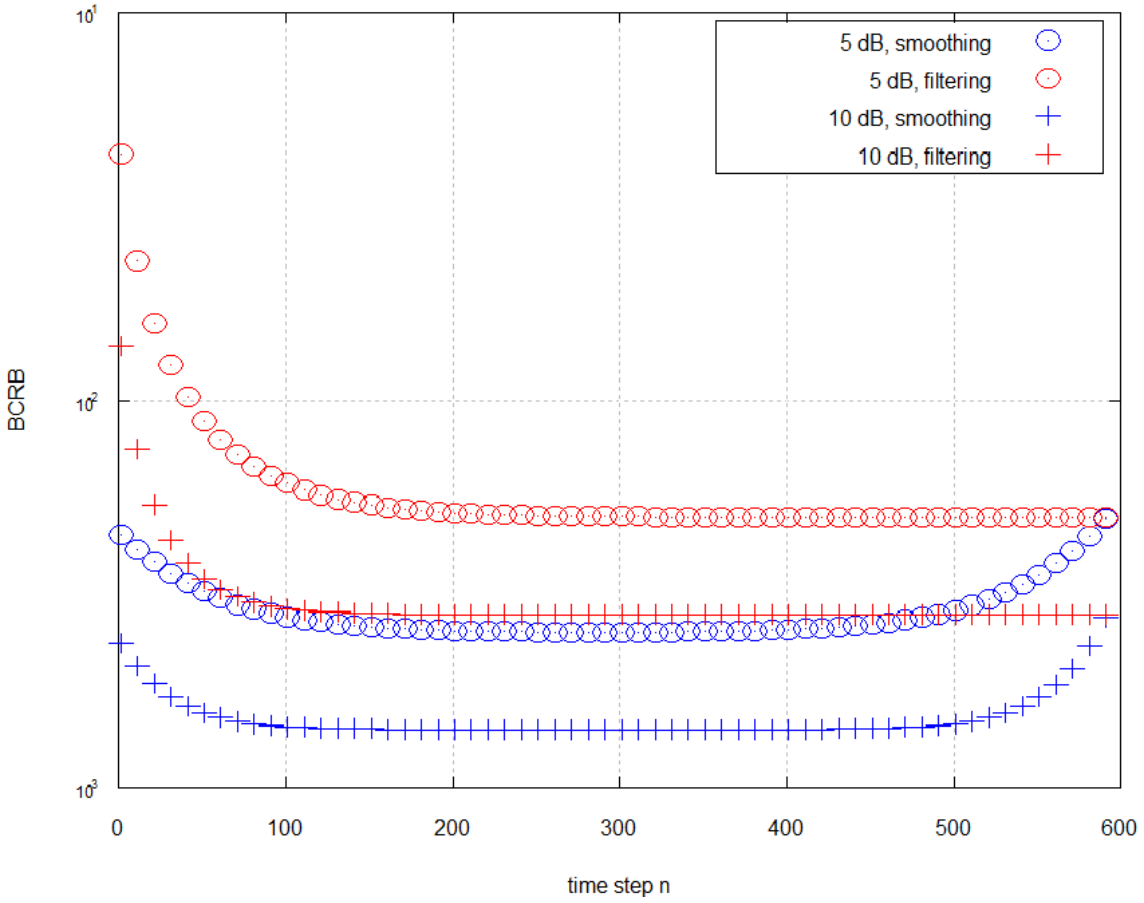

Figure 4.18. BCRB for 2×2 MIMO case as a function of the time step n.

Figure 4.19 shows the BCRBs and the performance for the two phase detectors in the 2×2 case. As it can be seen, both phase detectors approach the respective BCRB with sufficiently high SNR. The EM detector reaches within 1-2 dB from the BCRB, while about 3 dB separate the PLL solution from the BCRB for filtering.

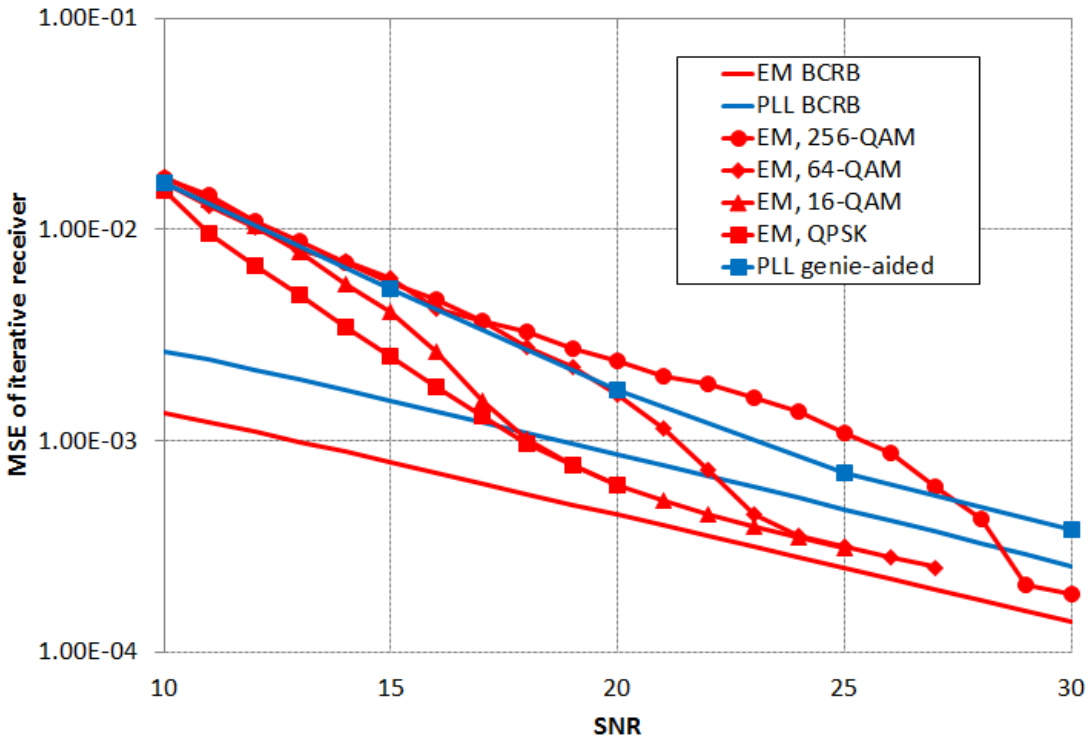

Figure 4.19. Performance of phase detectors. 2×2 MIMO case.

Chapter 5

Conclusions and acknowledgments

In this PhD thesis we wanted to make a step into the problem related to the improvement of wireless links affected by phase noise. We presented a low complex turbo synchronizer, feasible for actual application, achieving good performance close to the channel capacity. The evaluation of the capacity for multiple-anntennas channels has been also investigated, through a semi-analytical and non-asymptotic expression of a capacity upper bound and a simulation-based tool to compute a capacity lower bound, to give a complete set of results.

This work could be a basis for the design of other detection algorithms and for the evaluation of always tighter bounds to channel capacity.

I would like to thank Giuseppe Durisi, who helped me during my abroad period in Chalmers University in Gothenburg and Alberto Tarable, who has always supported me during the PhD program. I want to thank also Giulio Coluccia for his friendship and for giving me useful suggestions during my work. Last, but not the least, thank Guido Montorsi, a great mind and a pride for the Italian Academia.

Appendix A

Proof of Theorem 4

The asymptotic characterization (3.53) is obtained by proving that the upper bound (3.52) matches up to a $o(1)$ term the lower bound we shall report in Appendix A.1 below.

A.1 Lower bound

We take $\{\mathbf{x}_k\}$ i.i.d. and isotropically distributed according to Proposition 1. Specifically, we let $\mathbf{x}_k \sim \sqrt{\rho}\mathbf{z}_{\rho,\rho_0}$ where for a given $\rho_0 > 0$ the random variable $\mathbf{z}_{\rho,\rho_0}$ has the following pdf:

$$f_{\mathbf{z}_{\rho,\rho_0}}(\mathbf{a}) = \frac{f_{\mathbf{z}}(\mathbf{a})}{P\{\|\mathbf{z}\|^2 \geq \rho_0/\rho\}} \mathbb{1}\left\{\frac{\rho_0}{\rho} \leq \|\mathbf{a}\|^2 \leq 1\right\}. \tag{A.1}$$

Here, $\mathbb{1}\{\cdot\}$ denotes the indicator function and the pdf $f_{\mathbf{z}}$ of the random variable $\mathbf{z}$ is given by

$$f_{\mathbf{z}}(\mathbf{a}) = \left(M - \frac{1}{2}\right)\frac{\Gamma(M)}{\pi^M}\frac{1}{\|\mathbf{a}\|}\mathbb{1}\left\{\|\mathbf{a}\|^2 \leq 1\right\}. \tag{A.2}$$

Note that for every $\rho_0 > 0$ the pdf $f_{\mathbf{z}_{\rho,\rho_0}}$ converges pointwise to $f_{\mathbf{z}}$ as $\rho \to \infty$. The rationale behind the choice of $f_{\mathbf{z}}$ is that it turns out to maximize $h(\mathbf{z}) - \mathbb{E}[\log\|\mathbf{z}\|]$ under the constraint that $\|\mathbf{z} \leq 1$ w.p.1. The pdf $f_{\mathbf{z}_{\rho,\rho_0}}$ is constructed from $f_{\mathbf{z}}$ so as to guarantee that $\|\mathbf{x}_k\|^2 \geq \rho_0$, a property that will be useful in the remainder of the proof. To obtain the desired lower bound, we first proceed as in [**?**] and use chain rule for mutual information and that mutual information is non-negative to obtain

$$I(\mathbf{x}^n;\mathbf{y}^n) = \sum_{k=1}^{n} I(\mathbf{x}_k;\mathbf{y}^n|\mathbf{x}^{k-1}) \geq \sum_{k=2}^{n} I(\mathbf{x}_k;\mathbf{y}^k|\mathbf{x}^{k-1}). \tag{A.3}$$

Fix now $k \geq 2$ and set

$$\epsilon_k \triangleq I(\mathbf{x}_k;\theta_{k-1}|\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{x}_{k-1}). \tag{A.4}$$

We have

$$\begin{aligned} I(\mathbf{x}_k; \mathbf{y}^{|}\mathbf{x}^{k-1}) &\overset{(a)}{=} I(\mathbf{x}_k; \mathbf{y}^k, \mathbf{x}^{k-1}) \\ &\overset{(b)}{\geq} I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}) \\ &= I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}, \theta_{k-1}) - \epsilon_k \\ &\overset{(c)}{=} I(\mathbf{x}_k; \mathbf{y}_k, \theta_{k-1}) - \epsilon_k \\ &\overset{(d)}{=} I(\mathbf{x}_k; \mathbf{y}_k|\theta_{k-1}) - \epsilon_k \\ &\overset{(e)}{=} I(\mathbf{x}_2; \mathbf{y}_2|\theta_1) - \epsilon_2. \end{aligned} \tag{A.5}$$

Here, (a) follows because the $\{\mathbf{x}_k\}$ are independent; in (b) we used chain rule for mutual information and that mutual information is nonnegative; (c) follows because $\mathbf{x}_k$ and the pair $(\mathbf{y}_{k-1}, \mathbf{x}_{k-1})$ are conditionally independent given $(\theta_{k-1}, \mathbf{y}_k)$; (d) holds because $\mathbf{x}_k$ and θ_{k-1} are independent; finally (e) follows from stationarity.

Substituting(A.5) into (A.3) and then (A.3) into (3.38), we obtain

$$C(\rho) \geq I(\mathbf{x}_2; \mathbf{y}_2|\theta_1) - \epsilon_2. \tag{A.6}$$

We next investigate the two terms on the RHS of (A.6) separately. We shall show that the first term has the desired asymptotic expansion, while the second term can be made arbitrarily close to zero by choosing ρ_0 sufficiently large.

The first term on the RHS of (A.6) We write

$$I(\mathbf{x}_2; \mathbf{y}_2|\theta_1) = h(\mathbf{y}_2|\theta_1) - h(\mathbf{y}_2|\mathbf{x}_2, \theta_1) \tag{A.7}$$

and bound the two terms separately. For the first term, we have that

$$\begin{aligned} h(\mathbf{y}_2|\theta_1) &\geq h(\mathbf{y}_2|\mathbf{w}_2, \theta_1) \\ &= h(e^{j\theta_2}\mathbf{x}_2|\theta_1) \\ &\overset{(a)}{=} h(\mathbf{x}_2) \\ &= M \log \rho + h(\mathbf{z}_{\rho,\rho_0}) \\ &\overset{(b)}{=} M \log \rho - \log \frac{(M-1/2)\Gamma(M)}{\pi^M P\{\|\mathbf{z}\|^2 \geq \rho_0/\rho\}} + \mathbb{E}\left[\log \|\mathbf{z}_{\rho,\rho_0}\|\right]. \end{aligned} \tag{A.8}$$

Here (a) follows because $\mathbf{x}_2$ is isotropically distributed and (b) holds because of (A.1) and (A.2). For the second term on the RHS of (A.7), we proceed as follows. Let $\mathbf{x}_2 = s_2\mathbf{v}_2$, with $s_2 = \|\mathbf{x}_2\|$ and, hence, $s_2^2 \sim \rho\|\mathbf{z}_{\rho,\rho_0}\|^2$. Furthermore, let $z_2 \sim \mathcal{N}(0,1)$. Then, proceeding as in [24, Equation (10)]

$$h(\mathbf{y}_2|\mathbf{x}_2, \theta_1) = h(\mathbf{y}|s_2, \mathbf{v}_2, \theta_1) = h(e^{j\theta_2}s_2 + z_2|s_2, \theta_1) + \log(\pi e)^{M-1}. \tag{A.9}$$

The first term on the RHS of (A.9) can be bounded as follows

$$\begin{aligned} h(e^{j\theta_2}s_2+z_2|s_2,\theta_1) &= h(e^{j\theta_2}(s_2+z_2)|s_2,\theta_1) \\ &\overset{(a)}{=} h(e^{j\Delta}(s_2+z_2)|s_2) \\ &\overset{(b)}{=} h(|s_2+z_2|^2|s_2)+h(\phi_2(s_2^2)+\Delta||s_2+z_2|,s_2)-\log 2 \\ &\overset{(c)}{\leq} \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1+2\rho\|\mathbf{z}_{\rho,\rho_0}\|^2\right]\right)\right]+h(\phi_2(s_2^2)+\Delta|s_2)-\log 2. \end{aligned} \tag{A.10}$$

Here, in (a) we used (1.1) and denoted by Δ a random variable distributed as in (1.3); in (b) we evaluated the differential entropy in polar coordinates using [26, Lemma 6.15 and Lemma 6.16]. Finally, (c) follows because the Gaussian distribution maximizes differential entropy under a variance constraint and because conditioning reduces entropy. Note that

$$h(\phi_2(s_2^2)+\Delta|s_2) \leq \max_{\xi\geq\sqrt{\rho_0}} h(\phi_2(\xi^2)+\Delta) = h(\phi_2(\rho_0)+\Delta). \tag{A.11}$$

This term can be made arbitrarily close to $h(\Delta)$ by choosing ρ_0 in (A.1) sufficiently large. Summarizing, we have shown that

$$I(\mathbf{x}_2;\mathbf{y}_2|\theta_1) \geq M\log\rho - \log\frac{(M-1/2)\Gamma(M)}{\pi^M P\{\|\mathbf{z}\|^2\geq\rho_0/\rho\}} + \mathbb{E}\left[\log\|\mathbf{z}_{\rho,\rho_0}\|\right] \tag{A.12}$$

$$\begin{aligned} &- \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1+2\rho\|\mathbf{z}_{\rho,\rho_0}\|^2\right]\right)\right] \\ &- h(\phi_2(\rho_0)+\Delta) + \log 2 - \log(\pi e)^{M-1} \\ &= \left(M-\frac{1}{2}\right)\log\rho - \log\left(M-\frac{1}{2}\right) - \log\Gamma(M) \\ &+ \frac{1}{2}\log\pi - \left(M-\frac{1}{2}\right) - h(\phi_2(\rho_0)+\Delta) + o(1). \end{aligned} \tag{A.13}$$

Here, the last step follows because

$$\mathbb{E}\left[\log(1+c\rho\|\mathbf{z}_{\rho,\rho_0}\|^2)\right] = \log(c\rho) + \mathbb{E}\log\|\mathbf{z}_{\rho,\rho_0}\|^2 + o(1) \tag{A.14}$$

for all $c>0$, and

$$\lim_{\rho\to\infty} P\{\|\mathbf{z}\|^2\geq\rho_0/\rho\} = 1. \tag{A.15}$$

The second term on the RHS of (A.6) Let $\mathbf{x}_1 = s_1\mathbf{v}_1$ and $z_1\sim\mathcal{N}(0,1)$. Proceeding similarly as in [26, Appendix IX], we obtain (see [2, Equation (25)])

$$I(\mathbf{x}_2;\theta_1|\mathbf{y}_2,\mathbf{y}_1,\mathbf{x}_1) = h(\theta_2|e^{j\theta_1}(\sqrt{\rho_0}+z_1)) - h(\theta_2|\theta_1). \tag{A.16}$$

As claimed, the RHS of (A.16) can be made arbitrarily close to zero by choosing ρ_0 in (A.1) sufficiently large.

A.2 Upper Bound

We exploit the property that the high-SNR behavior of $C(\rho)$ does not change if the support of the input distribution is constrained to lie outside a sphere of arbitrary radius. This result, known as *escape-to-infinity* property of the capacity-achieving input distribution [26, Definition 4.11], is formalized in the following lemma.

Lemma 2. *Fix an arbitrary $\xi_0 > 0$ and let $\mathcal{K}(\xi_0) = \{\mathbf{x} \in \mathcal{C}^M : \|\mathbf{x}\| \geq \xi_0\}$. Denote by $C^{(\xi_0)}(\rho)$ the capacity of the channel (3.36) when the input signal is subject to the peak-power constraint (3.37) and to the additional constraint that $\mathbf{x}_k \in \mathcal{K}(\xi_0)$ almost surely for all k. Then*

$$C(\rho) = C^{(\xi_0)}(\rho) + o(1), \quad \rho \to \infty \tag{A.17}$$

with $C(\rho)$ given in (3.38).

Proof. The lemma follows directly from [26, Theorem 8] and [26, Theorem 4.12]. □

Fix $\xi_0 > 0$. By proceeding as in (3.52), we obtain [1]

$$C^{(\xi_0)}(\rho) \leq \sup\big\{I(\mathbf{y};\mathbf{x})\big\} + \log(2\pi) - h(\Delta) \tag{A.18}$$

where, this time, the supremum is over all probability distributions on $\mathbf{x}$ that satisfy $\|\mathbf{x}\|^2 \in [\xi_0^2, \rho]$ w.p.1. We next upper-bound $I(\mathbf{y};\mathbf{x})$ by using duality as in (3.49), i.e., we exploit that

$$I(\mathbf{y};\mathbf{x}) \leq -\mathbb{E}\log q_{\mathbf{y}}(\mathbf{y})\| - h(\mathbf{y}|\mathbf{x}) \tag{A.19}$$

for every output distribution $q_{\mathbf{y}}(\mathbf{y})$. We choose a different $q_{\mathbf{y}}(\mathbf{y})$ than the one resulting in (3.48). Roughly speaking, we want $q_{\mathbf{y}}(\mathbf{y})$ to be the output distribution induced by the input distribution (A.1) we used for the lower bound. When constructing $q_{\mathbf{y}}(\mathbf{y})$, we shall ignore the additive noise over the support of the input distribution, and consider the effect of the additive noise only outside an ϵ-neighborhood of the set $\{\mathbf{x} \in \mathcal{C}^M : \|\mathbf{x}\|^2 \leq \rho\}$.[2] Specifically, we shall set $\mathbf{r} \triangleq \mathbf{y}/\sqrt{\rho}$, $\mathcal{S}_\epsilon \triangleq \{\mathbf{r} \in \mathcal{C}^M : \|\mathbf{r} - \mathbf{x}'\| \leq \epsilon \text{for some} \|\mathbf{x}'\| \leq 1\}$ and choose the following probability distribution for $\mathbf{r}$

$$q_{\mathbf{r}}(\mathbf{r}) = \begin{cases} \frac{(M-1/2)}{\pi^M K_{\rho,\epsilon}} \frac{\Gamma(M)}{\|\mathbf{r}\|}, & \text{if} \quad \mathbf{r} \in \mathcal{S}_\epsilon \\ \frac{\rho^M}{\pi^M K_{\rho,\epsilon}} e^{-\rho\|\mathbf{r}\|^2}, & \text{if} \quad \mathbf{r} \notin \mathcal{S}_\epsilon \end{cases} \tag{A.20}$$

[1] To keep notation compact, we write $\mathbf{x}_0$ simply as $\mathbf{x}$; same convention for $\mathbf{y}_0$.

[2] This choice is inspired by [38], where the rates achievable with dense constellations over an AWGN channel (no phase noise) are analyzed.

where

$$K_{\rho,\epsilon} = \underbrace{\int_{\mathbf{r}\in\mathcal{S}_\epsilon} \frac{(M-1/2)}{\pi^M}\frac{\Gamma(M)}{\|\mathbf{r}}\, d\mathbf{r}\|}_{\triangleq K_{\infty,\epsilon}} + \int_{\mathbf{r}\notin\mathcal{S}_\epsilon} \frac{\rho^M}{\pi^M} e^{-\rho\|\mathbf{r}\|^2} d\mathbf{r}. \tag{A.21}$$

This yields

$$-\mathbb{E}\left[\log q_{\mathbf{y}}(\mathbf{y})\right] = M\log\rho - \mathbb{E}\left[\log q_{\mathbf{r}}(\mathbf{r})\right] \tag{A.22}$$

where

$$\begin{aligned}-\mathbb{E}\left[\log q_{\mathbf{r}}(\mathbf{r})\right] = &-\left[\log\frac{(M-1/2)\Gamma(M)}{\pi^M K_{\rho,\epsilon}}\right] P\{\mathbf{r}\in\mathcal{S}_\epsilon\} \\ &+ \mathbb{E}\left[(\log\|\mathbf{r}\|)\mathbb{1}\{\mathbf{r}\in\mathcal{S}_\epsilon\}\right] - \left[\log\frac{\rho^M}{\pi^M K_{\rho,\epsilon}}\right] P\{\mathbf{r}\notin\mathcal{S}_\epsilon\} \\ &+ \mathbb{E}\left[\rho\|\mathbf{r}\|^2\mathbb{1}\{\mathbf{r}\notin\mathcal{S}_\epsilon\}\right].\end{aligned} \tag{A.23}$$

We next characterize each term on the RHS of (A.23) in the limit $\rho\to\infty$.

The first term By construction (see (A.21)), we have that

$$\lim_{\epsilon\to 0}\lim_{\rho\to\infty} K_{\rho,\epsilon} = \lim_{\epsilon\to 0} K_{\infty,\epsilon} = 1. \tag{A.24}$$

Furthermore, let $\mathbf{w}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_M)$. Then

$$\begin{aligned}P\{\mathbf{r}\notin\mathcal{S}_\epsilon\} &\overset{(a)}{=} P\{\|\mathbf{w}\|/\sqrt{\rho}\geq\epsilon\} \\ &\overset{(b)}{=} \frac{\Gamma(M,\epsilon^2\rho)}{\Gamma(M)} \\ &\overset{(c)}{=} \frac{(\epsilon^2\rho)^{M-1}e^{-\epsilon^2\rho}}{\Gamma(M)} + o(\rho^{M-1}e^{-\epsilon^2\rho}), \quad \rho\to\infty.\end{aligned} \tag{A.25}$$

Here, (a) follows because $\mathbf{r}\sim\mathbf{x}'+\mathbf{w}/\sqrt{\rho}$ for some $\|\mathbf{x}'\|\leq 1$; in (b), the function $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function [39, Equation 6.5.3]; finally,(c) follows from [39, Equation 6.5.32]. Using (A.25), we conclude that the first term on the RHS of (A.23) admits the following asymptotic expansion:

$$\left[\log\frac{(M-1/2)\Gamma(M)}{\pi^M K_{\rho,\epsilon}}\right]\Pr\{\mathbf{r}\in\mathcal{S}_\epsilon\} = \log\frac{(M-1/2)\Gamma(M)}{\pi^M K_{\infty,\epsilon}} + o(1). \tag{A.26}$$

Furthermore, (A.24) implies that $K_{\infty,\epsilon}$ can be made arbitrarily close to 1 by choosing ϵ sufficiently small.

The second term Note that

$$\mathbb{E}\left[(\log\|\mathbf{r}\|)\mathbb{1}\{\mathbf{r}\in\mathcal{S}_\epsilon\}\right]=\frac{1}{2}\mathbb{E}\left[(\log\|\mathbf{y}\|^2)\mathbb{1}\{\mathbf{y}/\sqrt{\rho}\in\mathcal{S}_\epsilon\}\right]-\frac{1}{2}\log\rho. \tag{A.27}$$

Assume without loss of generality that $\rho>1$. Then $(\mathbf{y}/\sqrt{\rho})\notin\mathcal{S}_\epsilon$ implies that $\|\mathbf{y}\|^2>\rho>1$. Hence, we conclude that $\log\|\mathbf{y}\|^2>0$ whenever $(\mathbf{y}/\sqrt{\rho})\notin\mathcal{S}_\epsilon$. As a consequence, we can upper-bound (A.27) by adding

$$\frac{1}{2}\mathbb{E}\left[(\log\|\mathbf{y}\|^2)\mathbb{1}\{\mathbf{y}/\sqrt{\rho}\notin\mathcal{S}_\epsilon\}\right] \tag{A.28}$$

and obtain

$$\mathbb{E}\left[(\log\|\mathbf{r}\|)\mathbb{1}\{\mathbf{r}\in\mathcal{S}_\epsilon\}\right]\leq\frac{1}{2}\mathbb{E}\left[\log\|\mathbf{y}\|^2\right]-\frac{1}{2}\log\rho. \tag{A.29}$$

The third term It follows from (A.25) that

$$\left[\log\frac{\rho^M}{\pi^M K_{\rho,\epsilon}}\right]P\{\mathbf{r}\notin\mathcal{S}_\epsilon\}=o(1). \tag{A.30}$$

The fourth term We have that

$$\begin{aligned}\mathbb{E}\left[\rho\|\mathbf{r}\|^2\mathbb{1}\{\mathbf{r}\notin\mathcal{S}_\epsilon\}\right]&\overset{(a)}{\leq}\rho\sqrt{\mathbb{E}\left[\|\mathbf{r}\|^4P\{\mathbf{r}\notin\mathcal{S}_\epsilon\}\right]}\\&\overset{(b)}{=}o(1).\end{aligned} \tag{A.31}$$

Here, (a) follows from Chaucy-Schwarz inequality and (b) follows from (A.25). We next substitute (A.26), (A.27), (A.30), and (A.31) into (A.23) and then (A.23) into (A.22) and obtain

$$-\mathbb{E}\left[\log q_{\mathbf{y}}(\mathbf{y})\right]=\left(M-\frac{1}{2}\right)\log\rho-\log\frac{(M-1/2)\Gamma(M)}{\pi^M K_{\infty,\epsilon}}+\frac{1}{2}\mathbb{E}\left[\log\|\mathbf{y}\|^2\right]+o(1). \tag{A.32}$$

Set now $s=\|\mathbf{x}\|$ and $z\sim\mathcal{N}(0,1)$. By proceeding as in [2, Equation (33)], we can rewrite the conditional differential entropy $h(\mathbf{y}|\mathbf{x})$ on the RHS of (A.19) as

$$h(\mathbf{y}|\mathbf{x})=h(|s+z|^2|s)+\log\pi^M+M-1. \tag{A.33}$$

Substituting (A.32) and (A.33) into (A.19) and using that

$$\|\mathbf{y}\|^2\sim|s+z_1|^2+\sum_{j=2}^{M}|z_j|^2 \tag{A.34}$$

where $z_j \sim \mathcal{N}(0,1)$, $j = 1, \ldots, M$, we obtain

$$\begin{aligned} I(\mathbf{y};\mathbf{x}) &\leq \left(M - \frac{1}{2}\right)\log\rho - \log\frac{(M-1/2)\Gamma(M)}{K_{\infty,\epsilon}} \\ &\quad + \frac{1}{2}\mathbb{E}\left[\log\left(|s+z_1|^2 + \sum_{j=2}^{M}|z_j|^2\right)\right] - h(|s+z|^2|s) - (M-1) + o(1) \\ &\leq \left(M - \frac{1}{2}\right)\log\rho - \log\frac{(M-1/2)\Gamma(M)}{K_{\infty,\epsilon}} - (M-1) \\ &\quad + \max_{\xi_0 \leq \xi \leq \sqrt{\rho}}\left\{\frac{1}{2}\mathbb{E}\left[\log\left(|\xi+z_1|^2 + \sum_{j=2}^{M}|z_j|^2\right)\right] - h(|\xi+z|^2)\right\} + o(1) \end{aligned} \tag{A.35}$$

Substituting (A.35) into (A.18) and using that

$$\lim_{\xi\to\infty}\left\{\frac{1}{2}\mathbb{E}\left[\log\left(|\xi+z_1|^2 + \sum_{j=2}^{M}|z_j|^2\right)\right] - h\left(|\xi+z|^2\right)\right\} = -\frac{1}{2}\log(4\pi e) \tag{A.36}$$

which follows by [21, Equation (9)] and by proceeding similarly to the proof of [26, Lemma 6.9], we conclude that we can make the bound on $C^{(\xi_0)}(\rho)$ just derived to be arbitrarily close to (3.53) in the high-SNR regime by choosing ϵ sufficiently small and ξ_0 sufficiently large.

Appendix B

Evaluation of BRCB

In the recent years, the ever increasing request for fast wireless communications has urged to design systems with high-throughput LoS backhaul links. To obtain the desired efficiencies, two main engineering solutions have been resorted to: multiantenna (MIMO) links and high-efficiency modulations. In comparison with single-antenna (SISO) links, MIMO LoS systems offer a relevant throughput increase but, at the same time, are more sensitive to phase noise, especially when design considerations impose to have different oscillators feeding each antenna at both sides.

In this Appendix, we show the computation of the Bayesian Cramér-Rao bound (BCRB) for a MIMO LoS system with independent Wiener phase noise at each antenna. The BCRB gives a lower bound to the mean-square error achievable by any phase detector, and represents thus a fundamental tool to evaluate the performance of phase detectors. More in particular, we concentrate on large blocklengths and derive the BCRB in the middle of the block. The BCRB for the MIMO case with distributed phase noise is already computed in [40], where however no explicit expression of the BCRB is presented in the large-blocklength regime. Moreover, we obtain analytic expressions for the particular MIMO scenario where the channel matrix entries have constant modulus. As in [40, 37], we distinguish the case where information can be gleaned both from the past and from the future (as it is in smoothing algorithms) from the case where only the past symbols can be used (like in filtering algorithms).

The structure of the Appendix is as follows: in Section B.1, we describe the system of interest; in Section B.2, we derive the BCRB for any phase-detection algorithm; in Section B.3, we particularize the BCRB for filtering algorithms.

Throughout the treatment, matrices and vectors are denoted with boldface upper- and lower-case letters, respectively. $\mathbf{I}_N$ represents an identity matrix of size N, while $\mathbf{0}_{M,N}$ and $\mathbf{1}_{M,N}$ denote an all-zero and an all-one matrix with size $M \times N$ (the subscripts can be dropped whenever the dimension can be inferred from the context). $\mathbf{A}^{\mathsf{T}}$, $\mathbf{A}^*$ and $\mathbf{A}^{\mathsf{H}}$ denote the transpose, conjugate and transpose conjugate of matrix

$\mathbf{A}$. $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. Finally, $\odot$ represents the Hadamard (elementwise) product.

B.1 System description

We consider an $N_t \times N_r$ MIMO LoS channel with independent phase noise at each antenna, whose input-output relationship at time $n = 1,2, \ldots, T$ is given by

$$\mathbf{y}[n] = \mathbf{\Phi}_R[n]\, \mathbf{H}\, \mathbf{\Phi}_T[n]\mathbf{x}[n] + \mathbf{z}[n] \tag{B.1}$$

where:

- $\mathbf{H}$ is the $N_r \times N_t$ LOS channel matrix, assumed to be constant and known at the receive side;
- $\mathbf{\Phi}_T[n] = \text{diag}\left(e^{j\phi_1[n]}, \ldots, e^{j\phi_{N_t}[n]}\right)$ and $\mathbf{\Phi}_R[n] = \text{diag}\left(e^{j\phi_{N_t+1}[n]}, \ldots, e^{j\phi_{N_t+N_r}[n]}\right)$ are the diagonal matrices of transmit and receive phase-noise coefficients at time n, respectively, assumed to be unknown at both sides;
- $\mathbf{x}[n]$ and $\mathbf{y}[n]$ are the column vectors of the N_t transmitted symbols and N_r received samples at time n, respectively;
- $\mathbf{z}[n]$ is a size-N_r vector of zero-mean, circularly-invariant Gaussian-noise samples, with variance σ^2 per real dimension, which are supposed to be independent across time and receive antenna.

Notice that the fact of having a fixed, known channel matrix arises from the hypothesis of LoS conditions [17].

We will suppose hereafter that the transmitted symbols are i.i.d., with zero mean and normalized average power $\mathbb{E}|x_i[n]|^2 = 1$, where $x_i[n]$ is the i-th element of vector $\mathbf{x}[n]$

For the phase-noise samples, time dependency is kept into account by assuming Wiener phase-noise processes:

$$\phi_i[n] = \phi_i[n-1] + w_i[n], \;\; i = 1, \ldots, N_r + N_t,\; n = 1,2, \ldots \tag{B.2}$$

where $\phi_1[0], \ldots, \phi_{N_r+N_t}[0]$ are independent and uniformly distributed over $[0,2\pi)$ and $w_i[n], \ldots, w_{N_r+N_t}[n]$, are independent zero-mean white Gaussian processes with power ρ^2 (all processes have the same power, unlike in [40]). Each tap of the MIMO channel is affected by the sum of one transmit and one receive phase-noise process. Define a *sum* phase-noise process as:

$$\phi_{ii'}[n] = \phi_i[n] + \phi_{N_t+i'}[n], \;\; i = 1, \ldots, N_t,\; i' = 1, \ldots, N_r \tag{B.3}$$

Unlike the *atomic* $\phi_i[n]$ processes, sum phase-noise processes are correlated. Indeed, they can all be written as linear functions of the basis $\{\{\phi_{i1}[n],\ i = 1, \ldots, N_t\}, \{\phi_{1i}[n],\ i = 2, \ldots, N_r\}\}$. Although atomic phase-noise processes have a simpler statistical characterization, they are not observable, so that phase estimation must pass through sum phase-noise process estimation.

We define for future use the size-$(N_t + N_r)$ vector $\boldsymbol{\phi}[n]$, whose i-th element is $\phi_i[n]$, $i = 1, \ldots, N_r + N_t$, and the size-$(N_t + N_r - 1)$ vector $\boldsymbol{\phi}^{\text{sum}}[n]$, which contains the elements of the basis defined above. Furthermore, $\bar{\boldsymbol{\phi}} = [\boldsymbol{\phi}[1], \ldots, \boldsymbol{\phi}[T]]$ and $\bar{\boldsymbol{\phi}}^{\text{sum}} = [\boldsymbol{\phi}^{\text{sum}}[1], \ldots, \boldsymbol{\phi}^{\text{sum}}[T]]$.

B.2 Derivation of the BCRB

In this section, we compute the BCRB for the system described in the previous section, in the hypothesis that the transmitted symbols are *known* at the receiver. This hypothesis well suits the case of an iterative receiver based on the "turbo" principle, which iterates between phase detector and channel decoder for a certain number of times: if convergence eventually occurs, the output of the channel decoder after several iterations provides the phase detector with (almost) perfect knowledge of the transmitted symbols. If pilots are inserted into the transmitted symbol sequence, the analysis below can also be applied to a rough phase estimate based on such pilots only.

The BCRB allows to lower-bound the mean-square error (MSE) between the unknown phase-noise sample sequence and its estimate performed by the phase detector at the receiver. Let $\widehat{\bar{\boldsymbol{\phi}}}^{\text{sum}}$ be any possible estimate of $\bar{\boldsymbol{\phi}}^{\text{sum}}$. The covariance matrix $\boldsymbol{\Sigma}$ of such estimate is given by

$$\boldsymbol{\Sigma} = \mathbb{E}_{\bar{\boldsymbol{\phi}},\bar{\mathbf{y}},\bar{\mathbf{x}}} \left\{ \left(\widehat{\bar{\boldsymbol{\phi}}}^{\text{sum}} - \bar{\boldsymbol{\phi}}^{\text{sum}} \right) \left(\widehat{\bar{\boldsymbol{\phi}}}^{\text{sum}} - \bar{\boldsymbol{\phi}}^{\text{sum}} \right)^{\mathsf{T}} \right\} \tag{B.4}$$

The BCRB then states that

$$\boldsymbol{\Sigma} \succeq \mathbf{M}^{-1} \tag{B.5}$$

where $\mathbf{M}$ is the Bayesian information matrix (BIM) defined by

$$\mathbf{M} = -\mathbb{E}_{\bar{\boldsymbol{\phi}},\bar{\mathbf{y}},\bar{\mathbf{x}}} \left\{ \nabla_{\bar{\boldsymbol{\phi}}^{\text{sum}}} \nabla_{\bar{\boldsymbol{\phi}}^{\text{sum}}}{}^{\mathsf{T}} \log f\left(\bar{\mathbf{y}}, \bar{\boldsymbol{\phi}}^{\text{sum}} | \bar{\mathbf{x}}\right) \right\}. \tag{B.6}$$

Conditioning on $\bar{\mathbf{x}}$ in the above definition corresponds to the hypothesis of known transmitted symbols. The BCRB implies that:

$$(\boldsymbol{\Sigma})_{i,i} \geq (\mathbf{M}^{-1})_{i,i} \tag{B.7}$$

I.e., the BCRB provides lower bounds to the MSE for every estimator of the sum phase-noise samples.

The computation of the BIM has been carried out in [**?**] for a system very similar to the one in (B.1), and, in a more general setting, in [**?**]. The result is summarized in the following proposition.

Proposition 3. *For the channel model in* (B.1)*, if* $\rho \ll 2\pi$*, the BIM* $\mathbf{M}$ *defined in* (B.6) *is given by*

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_1 & & & \\ \mathbf{M}_1 & \mathbf{M}_0 & \mathbf{M}_1 & & \\ & \mathbf{M}_1 & \mathbf{M}_0 & \ddots & \\ & & \ddots & \ddots & \mathbf{M}_1 \\ & & & \mathbf{M}_1 & \mathbf{M}_0 \end{bmatrix} \tag{B.8}$$

where $\mathbf{M}_0 = \mathbf{J}^{\mathsf{T}}\widetilde{\mathbf{M}}_0\mathbf{J}$ *and* $\mathbf{M}_1 = \mathbf{J}^{\mathsf{T}}\widetilde{\mathbf{M}}_1\mathbf{J}$*, being* $\mathbf{J}$ *the Jacobian of the transformation from sum processes to atomic processes at a given time* n*,* $\widetilde{\mathbf{M}}_1 = -\frac{1}{\rho^2}\mathbf{I}_{N_t+N_r}$ *and* $\widetilde{\mathbf{M}}_0 = \widetilde{\mathbf{M}}_0^Y + \frac{2}{\rho^2}\mathbf{I}_{N_t+N_r}$*, and finally*

$$\widetilde{\mathbf{M}}_0^Y = \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{\Gamma}_T & (\mathbf{H}\odot\mathbf{H}^*)^{\mathsf{T}} \\ \mathbf{H}\odot\mathbf{H}^* & \mathbf{\Gamma}_R \end{bmatrix} \tag{B.9}$$

being $\mathbf{\Gamma}_T$ *and* $\mathbf{\Gamma}_R$ *two diagonal matrices whose diagonal elements are given by* $\mathbf{e}_i\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{e}_i$*,* $i = 1, \ldots, N_t$*, and* $\mathbf{e}_i\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_i$*,* $i = 1, \ldots, N_r$*, respectively.*

Proof: The computation of the BIM $\mathbf{M}$ can be performed by computing first the BIM for the atomic processes, which is given by

$$\widetilde{\mathbf{M}} = -\mathbb{E}_{\bar{\boldsymbol{\phi}},\bar{\mathbf{y}},\bar{\mathbf{x}}}\left\{\nabla_{\bar{\boldsymbol{\phi}}}\nabla_{\bar{\boldsymbol{\phi}}}{}^{\mathsf{T}}\log f\left(\bar{\mathbf{y}},\bar{\boldsymbol{\phi}}|\bar{\mathbf{x}}\right)\right\}. \tag{B.10}$$

and applying the following relationship between BIMs:

$$\mathbf{M} = \overline{\mathbf{J}}^{\mathsf{T}}\widetilde{\mathbf{M}}\overline{\mathbf{J}} \tag{B.11}$$

where

$$\overline{\mathbf{J}} = \mathrm{diag}\left(\mathbf{J}, \mathbf{J}, \ldots, \mathbf{J}\right) \tag{B.12}$$

The hypothesis $\rho^2 \ll 2\pi$, which is usually realistic, allows to approximate the Wiener model as

$$f(\boldsymbol{\phi}[n+1]|\boldsymbol{\phi}[n]) \simeq K' \prod_{i=1}^{N_t+N_r} e^{-\frac{(\phi_i[n+1]-\phi_i[n]+2\pi k_i[n])^2}{2\rho^2}} \tag{B.13}$$

where K' is a constant and $k_i[n]$ is a signed integer that minimizes the modulus of the exponent.

For more details, see [40].

■

The inversion of matrix $\mathbf{M}$ can be performed by using standard results on the inversion of tridiagonal symmetric matrices. Let $\mathbf{M}_{n,n}^{\dagger}$ be the n-th size-(N_t+N_r-1) square diagonal block of the inverse matrix $\mathbf{M}^{-1}$. It satisfies

$$\mathbf{M}_{n,n}^{\dagger} = (\mathbf{M}_0^Y + \mathbf{\Xi}[n] + \mathbf{\Theta}[n])^{-1} \tag{B.14}$$

where $\mathbf{M}_0^Y = \mathbf{J}^{\mathsf{T}}\widetilde{\mathbf{M}}_0^Y\mathbf{J}$. The matrices $\mathbf{\Xi}[n]$ and $\mathbf{\Theta}[n]$ are computed through forward and backward recursions as

$$\mathbf{\Xi}[n] = -\mathbf{M}_1 - \mathbf{M}_1\left(\mathbf{M}_0^Y - \mathbf{M}_1 + \mathbf{\Xi}[n-1]\right)^{-1}\mathbf{M}_1, \tag{B.15}$$

$$\mathbf{\Theta}[n] = -\mathbf{M}_1 - \mathbf{M}_1\left(\mathbf{M}_0^Y - \mathbf{M}_1 + \mathbf{\Theta}[n+1]\right)^{-1}\mathbf{M}_1 \tag{B.16}$$

with initial conditions $\mathbf{\Xi}[0]$ and $\mathbf{\Theta}[T]$.

In this paper, we are interested in the large-blocklength regime, so that we let $T \to \infty$. With this hypothesis, $\mathbf{\Xi}[T/2] = \mathbf{\Xi}[\infty] = \lim_{n\to\infty}\mathbf{\Xi}[n]$. By defining

$$\widetilde{\mathbf{\Xi}}[\infty] = \mathbf{M}_1^{-1}\mathbf{\Xi}[\infty] \tag{B.17}$$

and

$$\mathbf{M}_{10} = \mathbf{M}_1^{-1}\mathbf{M}_0^Y. \tag{B.18}$$

we find that $\widetilde{\mathbf{\Xi}}[\infty]$ must satisfy the fixed-point equation

$$\widetilde{\mathbf{\Xi}}[\infty] = -\mathbf{I} - \left(\mathbf{M}_{10} - \mathbf{I} + \widetilde{\mathbf{\Xi}}[\infty]\right)^{-1}. \tag{B.19}$$

Now, if the diagonalization of $\mathbf{M}_{10}$ reads as

$$\mathbf{M}_{10} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^{-1} \tag{B.20}$$

then we can verify that $\widetilde{\mathbf{\Xi}}[\infty]$ satisfies

$$\widetilde{\mathbf{\Xi}}[\infty] = \mathbf{\Psi}\mathbf{\Delta}\mathbf{\Psi}^{-1} \tag{B.21}$$

with the diagonal matrix $\mathbf{\Delta}$ given by

$$\mathbf{\Delta} = -\frac{\mathbf{\Lambda}}{2} - \left(\left(\frac{\mathbf{\Lambda}}{2}\right)^2 - \mathbf{\Lambda}\right)^{1/2} \tag{B.22}$$

irrespective of the initial conditions. The same result can be obtained for $\mathbf{\Theta}[T/2]$.

Substituting the above expressions into (B.14), we then find

$$\mathbf{M}_{T/2,T/2}^{\dagger} = -\mathbf{\Psi}\left(\mathbf{\Lambda}^2 - 4\mathbf{\Lambda}\right)^{-1/2}\mathbf{\Psi}^{-1}\mathbf{M}_1^{-1} \tag{B.23}$$

We can go on with computations for the case of a constant-modulus channel matrix, where each element of $\mathbf{H}$ has the same modulus α. This case is practically important when the distance between the transmit antenna array and the receive antenna array is large with respect to antenna separation at both sides [17]. In such a case, the MSE in the middle of the block will be the same for all sum processes by symmetry. The straightforward but tedious computations give

$$\begin{aligned}(\mathbf{M}^{-1})_{T/2,T/2} =& a\left(\frac{1}{N_t}+\frac{1}{N_r}\right)+b\left(1-\frac{1}{N_t}\right)\\ &+c\left(1-\frac{1}{N_r}\right)\end{aligned} \tag{B.24}$$

where

$$a = \frac{\sigma}{\alpha}\left(N_t+N_r\right)^{-1/2}\left(\frac{\alpha^2}{\sigma^2}(N_t+N_r)+\frac{4}{\rho^2}\right)^{-1/2} \tag{B.25}$$

$$b = \frac{\sigma}{\alpha}N_r^{-1/2}\left(\frac{\alpha^2}{\sigma^2}N_r+\frac{4}{\rho^2}\right)^{-1/2} \tag{B.26}$$

$$c = \frac{\sigma}{\alpha}N_t^{-1/2}\left(\frac{\alpha^2}{\sigma^2}N_t+\frac{4}{\rho^2}\right)^{-1/2} \tag{B.27}$$

B.3 The BCRB for filtering

The above computed BCRB holds for smoothing algorithms that use both past and future samples to compute the phase estimates. Instead, filtering algorithms (like the classical PLL algorithm) only consider past phase samples. Because of that, the BCRB must be changed in order to take into account this constraint.

When filtering is considered, only the forward recursion is to be considered. Thus (B.14) becomes

$$(\mathbf{M}^{\dagger}_{n,n})^{\text{filt}} = (\mathbf{M}_0^Y + \mathbf{\Xi}[n])^{-1} \tag{B.28}$$

where $\mathbf{\Xi}[n]$ obeys the same recursion (B.15).

In the case of a constant-modulus $\mathbf{H}$ with filtering, the BCRB retains the same expression as in (B.24), but with different parameters:

$$a^{\text{filt}} = \left(\frac{\alpha^2}{\sigma^2}\frac{N_t+N_r}{2}+\frac{a^{-1}}{2}\right)^{-1} \tag{B.29}$$

$$b^{\text{filt}} = \left(\frac{\alpha^2}{\sigma^2}\frac{N_r}{2}+\frac{b^{-1}}{2}\right)^{-1} \tag{B.30}$$

$$c^{\text{filt}} = \left(\frac{\alpha^2}{\sigma^2}\frac{N_t}{2}+\frac{c^{-1}}{2}\right)^{-1} \tag{B.31}$$

where a, b and c are given in (B.25)-(B.27).

Bibliography

[1] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *Communications, IEEE Transactions on*, vol. 59, no. 12, pp. 3223–3228, 2011.

[2] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of mimo wiener phase-noise channels," in *Information Theory and Applications Workshop (ITA), 2013*. IEEE, 2013, pp. 1–7.

[3] G. Durisi, A. Tarable, and T. Koch, "On the multiplexing gain of mimo microwave backhaul links affected by phase noise," in *Proc. IEEE Int. Conf. Commun.(ICC), Budapest, Hungary, Jun. 2013.*, 2013.

[4] M. Coldrey, H. Koorapaty, J. Berg, Z. Ghebretensae, J. Hansryd, A. Derneryd, and S. Falahati, "Small-cell wireless backhauling: A non-line-of-sight approach for point-to-point microwave links," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*. IEEE, 2012, pp. 1–5.

[5] J. Hansryd and J. Edstam, "Microwave capacity evolution," *The data boom: opportunities and challenges*, p. 22, 2011.

[6] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in mimo systems," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4790–4807, 2012.

[7] S. Chinnici, C. Decanis, A. Quadrini, and D. Weinholt, "High order m-qam transceiver for gigabit radio microwave transmission: Fpga test chipset and asic design," in *New Circuits and Systems Conference (NEWCAS), 2011 IEEE 9th International*. IEEE, 2011, pp. 297–300.

[8] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An em-based phase-noise estimator for mimo systems," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3215–3219.

[9] G. Colavolpe, "Communications over phase-noise channels: A tutorial review," in *Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), 2012 6th*. IEEE, 2012, pp. 316–327.

[10] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase

noise on uplink performance of large mu-mimo systems," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1190–1197.

[11] U. Mengali and A. N. D'Andrea, *Synchronization techniques for digital receivers*. Springer, 1997.

[12] A. Cero, "Advanced modulation/demodulation schemes for wireless communications," Ph.D. dissertation, University of Parma, 2009.

[13] H. Ghozlan and G. Kramer, "On wiener phase noise channels at high signal-to-noise ratio," *arXiv preprint arXiv:1301.6923*, 2013.

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[15] P. F. Driessen and G. J. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: A geometric interpretation," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 173–176, 1999.

[16] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor mimo wireless channels: Models and performance prediction," *Communications, IEEE Transactions on*, vol. 50, no. 12, pp. 1926–1934, 2002.

[17] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank line-of-sight mimo channels," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 4, pp. 1420–1425, 2007.

[18] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time wiener's phase noise channel," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1480–1486, 2012.

[19] D.-M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *Information Theory, IEEE Transactions on*, vol. 52, no. 8, pp. 3498–3508, 2006.

[20] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 284–287, 1974.

[21] A. Lapidoth, "On phase noise channels at high snr," in *Information Theory Workshop, 2002. Proceedings of the 2002 IEEE*. IEEE, 2002, pp. 1–4.

[22] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent awgn channel," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 866–883, 2005.

[23] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent awgn channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2257–2270, 2004.

[24] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *Communications Letters, IEEE*, vol. 16, no. 8, pp. 1157–1160, 2012.

[25] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *Information Theory, IEEE Transactions on*, vol. 55, no. 6, pp. 2716–2755, 2009.

[26] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2426–2467, 2003.

[27] G. Wunder, R. F. Fischer, H. Boche, S. Litsyn, and J.-S. No, "The papr problem in ofdm transmission: New directions for a long-lasting problem," *arXiv preprint arXiv:1212.2865*, 2012.

[28] J. G. Smith, "The information capacity of amplitude-and variance-constrained sclar gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[29] A. Lapidoth, "On the asymptotic capacity of stationary gaussian fading channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 437–446, 2005.

[30] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 437–453, 2006.

[31] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 9, pp. 1748–1757, 2005.

[32] J. Dauwels and H.-A. Loeliger, "Phase estimation by message passing," in *Communications, 2004 IEEE International Conference on*, vol. 1. IEEE, 2004, pp. 523–527.

[33] J. Dauwels, S. Korl, and H.-A. Loeliger, "Expectation maximization for joint decoding and phase estimation."

[34] N. Hadaschik, M. Dorpinghaus, A. Senst, O. Harmjanz, U. Kaufer, G. Ascheid, and H. Meyr, "Improving mimo phase noise estimation by exploiting spatial correlations," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 3. IEEE, 2005, pp. iii–833.

[35] H. Mehrpouyan, A. A. Nasir, T. Eriksson, S. D. Blostein, G. K. Karagiannidis, and T. Svensson, "Time-varying phase noise and channel estimation in mimo systems," in *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2012. Cesme, 17-20 June 2012*, 2012, pp. 560–564.

[36] A. Tarable, C. Camarda, and G. Montorsi, "Cramer-rao bounds for mimo los systems affected by distributed wiener phase noise in the large-blocklength regime," 2014.

[37] J. Dauwels, "Computing bayesian cramer-rao bounds," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on.* IEEE, 2005, pp.

425–429.

[38] T. Koch, A. Martinez *et al.*, "The capacity loss of dense constellations," *arXiv preprint arXiv:1205.5522*, 2012.

[39] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Dover Publications, 2012.

[40] A. A. Nasir, H. Mehrpouyan, R. Schober, and Y. Hua, "Phase noise in mimo systems: Bayesian cramer-rao bounds and soft-input estimation," 2013.